

'Der elektronische Zugriff auf die Quellen': zum Entwicklungsstand digitaler Bestandserschließung

Sahle, Patrick

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Sahle, P. (1999). 'Der elektronische Zugriff auf die Quellen': zum Entwicklungsstand digitaler Bestandserschließung. *Historical Social Research*, 24(2), 99-134. <https://doi.org/10.12759/hsr.24.1999.2.99-134>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

„Der elektronische Zugriff auf die Quellen“ Zum Entwicklungsstand digitaler Bestandserschließung

*Patrick Sahle**

Abstract: Libraries, archives and other institutions hold large collections of historical documents of all type and relevant to the humanities. The electronic accessibility of these holdings offers wide new research opportunities. The article summarizes a study on 75 research-projects (mainly located in Germany) aiming on new forms of availability of digital collections. It tries to give an overview on the main trends in the field of accessing data through. Furthermore a rough model for the whole process of developing these holdings is proposed. This consists of these parts: the representation (by digitization), the description (by metadata) and the publication (via user-interfaces) of the material. The single steps of the process are described and evaluated in their different state of development.

Vorbemerkungen

Auf dem Weg in die Informationsgesellschaft besteht in Bibliotheken, Archiven und anderen Institutionen, die Dokumenten- und Quellenbestände aller Art besitzen, die Tendenz zur Erschließung dieser Bestände unter den gegebenen neuen technischen Möglichkeiten. Mit dem Wandel der Werkzeuge gehen dabei grundsätzlichere Veränderungen einher, die alle Bereiche solcher Erschließungsvorhaben betreffen. Es wird nicht die gleiche Arbeit mit anderen Mitteln geleistet, vielmehr ist ein fundamentaler Wandel zu beobachten, der eine Reihe von Fragen ganz neu aufwirft: welche Bestände sollen überhaupt erschlossen werden, warum sollen sie erschlossen werden, in welcher Art und Weise und in welcher Tiefe, in welcher Form sollen sie den Benutzern

* Address all communications to Patrick Sahle, Riehler Str. 225, D-50735 Köln. [E-Mail: Sahle@uni-koeln.de](mailto:Sahle@uni-koeln.de) or confer for further addresses the personal homepage at <http://www.uni-koeln.de/~ahz26/>

zugänglich gemacht werden und wie ist dies schließlich alles zu organisieren und durchzuführen?

Es handelt sich bei der digitalen Bestandserschließung um ein wissenschaftliches Arbeitsgebiet, das sich in einem Stadium grundlegender Umbrüche befindet. So kann es nicht verwundern, daß es einerseits bereits eine große Zahl von Pilotprojekten gibt, die das Feld in seinen Möglichkeiten ausloten, daß aber andererseits auch ein hohes Maß an Orientierungslosigkeit, Verunsicherung sowie Informations- und Kommunikationsmängeln zu konstatieren sind. Orientierungslosigkeit resultiert aus mangelnden „Vorbildern“, nicht vorhandenem theoretischem Fundament, fehlenden Richtlinien und der Unbekanntheit der Anforderungen späterer Benutzer. Verunsicherung ergibt sich oft aus der ungeklärten rechtlichen Lage, ebenso aber aus der rasanten technischen Entwicklung, welche die dauerhafte Nutzbarkeit und Archivierbarkeit aller jetzt geleisteten Arbeit fraglich erscheinen läßt. Informations- und Kommunikationsmängel schließlich entstehen durch die breite Streuung der einzelnen Projekte auf die verschiedensten Arten von Institutionen und Fachgebieten und durch das hohe Maß an notwendiger interdisziplinärer Zusammenarbeit. Hier treffen oft hoch spezialisierte Fachleute aufeinander, denen ohne eine gemeinsame Gesamtperspektive die Blickwinkel der jeweils anderen verschlossen bleiben.

Gleichzeitig ist zu erwarten, daß die Pilotprojekte der Formierungsphase für die Zukunft Leitbilder mit hoher normierender Kraft sein werden. Dies verstärkt die Notwendigkeit einer frühzeitigen Reflexion über den ganzen Bereich. Eine am Kölner Zentralarchiv für empirische Sozialforschung, bzw. dem Zentrum für Historische Sozialforschung von November 1998 bis März 1999 durchgeführte Untersuchung sollte dazu einen Beitrag zur Gewinnung eines besseren Überblicks über das Forschungsgebiet „digitaler Bestandserschließung“ leisten und die dort vorhandenen hauptsächlichlichen Problemstellungen und Entwicklungstendenzen herausarbeiten. Mittelfristig sollten Information, Orientierung und Kommunikation in diesem Bereich verbessert werden. Langfristig muß es darum gehen, dem Feld der digitalen Bestandserschließung durch Abstraktion, Begriffsbildung, Strukturierung und Segmentierung ein referenzfähiges theoretisches und methodisches Gerüst zu geben. Die gesamte Dokumentation der Studie ist im Internet einzusehen, der vorliegende Beitrag ist eine leicht veränderte und gekürzte Fassung des ebenfalls dort einzusehenden Abschlußberichtes.¹

¹ Die Adresse lautet <http://www.uni-koeln.de/~ahz26/zhsf/>. Die Dokumentation enthält neben einer ausführlicheren Zusammenfassung (in HTML und Word97 bzw. Word6.0/95) Projektlisten, geordnet nach Bestandsart und geographischer Zuordnung, Auswahllisten zu Browsing-Konzepten und Suchmasken, ein Glossar und weitere Verweise zum Thema. Der dauerhafte Bestand der o.g. Adresse kann ebenso wenig garantiert werden, wie die fortlaufende Aktualisierung der Dokumentation. Man bediene sich deshalb bei nicht mehr gültigen Angaben und Verweisen entsprechender Suchmaschinen, beginnend mit dem Stichwort „Der elektronische Zugriff auf die Quellen“ bzw. „Vergleichende Studie zu

Der Untersuchungsbereich umfaßte Projekte zur „digitalen Bestandserschließung“. Unter „digital“ sollte dabei die Anwendung von Techniken verstanden werden, durch die es zu einer Art von Erschließung oder Repräsentation kommt, die mit herkömmlichen Mitteln (Druck, Karteisysteme, etc.) nicht zu realisieren wäre. Dies kann die digitale Abbildung von Dokumenten oder auch die elektronische Verwaltung von Erschließungsinformationen betreffen. Der Begriff „Bestand“ wurde unabhängig von den speziellen Definitionen, die im Bibliotheks- bzw. im Archivwesen Verwendung finden, sehr allgemein gefaßt, als größere Menge an Dokumenten bzw. Datenträgern, die Informationen enthalten, welche für die geistes- oder sozialwissenschaftliche Forschung von Interesse sind.² Beim Begriff der „Erschließung“ wurde zur Abgrenzung des Untersuchungsbereiches von einer „Mindesterschließungstiefe“ ausgegangen, die über die Digitalisierung (digitale Abbildung), die grobe Beschreibung des Gesamtbestandes oder aber die reine (traditionelle) Katalogisierung hinausgeht.

Gegenstand der Untersuchung waren schließlich 75 Projekte.³ Der Schwerpunkt der Untersuchung lag mit 61 Projekten im deutschsprachigen Bereich, während das übrige Europa mit 8 und die außereuropäische Welt mit 5 Unternehmungen vertreten waren. Die Erfassung der Projekte kann weder als vollständig, noch als repräsentativ angesehen werden.⁴ Die Gesamtzahl von 61 Projekten im Kernbereich der Untersuchung dürfte aber groß genug sein, um Tendenzaussagen zuzulassen. Die 14 weiteren Projekte dienten als Kontrollgruppe, um ggf. wesentliche Unterschiede zum internationalen Kontext feststellen zu können.

Projekten zur digitalen Bestandserschließung und zur Erschließung digitaler Bestände“. Dies dürfte auch langfristig eindeutig sein.

² Im Regelfall sind solche Bestände materielle Sammlungen in Bibliotheken, Archiven, Universitäts- oder Forschungsinstituten. Neben solchen „realen“ Beständen wurden aber auch jene mit in die Untersuchung einbegriffen, die als thematischer Bestand nicht an eine bestimmte Institution gebunden sind oder in einer solchen Institution „quer“ zur eigentlichen Bestandsordnung verlaufen. Dies mag dem allgemeinen Begriffsverständnis widersprechen - eine Einbeziehung auch thematisch orientierter Projekte ohne zugrundeliegenden realen Bestand scheint aber sinnvoll, weil die konzeptionellen Ansätze und die verwendeten Techniken sich nicht von anderen Projekten unterscheiden. Grundsätzlich wären auch Gruppen von „digital born documents“, die nicht durch ein bestimmtes materielles Trägermedium geprägt sind, mit einzubeziehen gewesen. Es ließ sich aber kein entsprechendes Projekt finden.

³ In den Projektweisern der Internetdokumentation können sich andere Summen ergeben, weil für eine gewisse Zeit noch Projekte hinzugefügt oder entfernt werden könnten. In der Untersuchung wurde der Datendichte kein hohes Gewicht gegeben. Es sollte deshalb nicht verwundern, wenn sich Auszählungen auf andere Basiswerte beziehen.

⁴ Da Vollständigkeit zumindest angestrebt war, spielten Überlegungen zur Repräsentativität keine Rolle. Die Existenz der meisten Projekte verdankt sich außerdem äußeren Faktoren, nämlich speziellen Förderprogrammen der Volkswagen-Stiftung bzw. der Deutschen Forschungsgemeinschaft, so daß ohnehin keine Folie einer „normalen“ Entwicklung möglich ist, vor der die Frage nach Repräsentativität sinnvoll zu stellen wäre.

Die Informationserhebung über die Projekte wurde mit einem datenbankgestützten offenen Frageraster durchgeführt, das einerseits dem Versuch diente, Kategorien zu finden, die sich - unabhängig von deren konkreter Formierung - auf alle Erschließungsprojekte anwenden lassen, so daß ein verallgemeinernder Vergleich möglich wurde. Andererseits ermöglichte es die gezielte Nachfrage in Detailbereichen, um Besonderheiten oder spezielle Probleme und Lösungen mit in den Überblick einfließen lassen zu können. Grundsätzlich wurde eine Einzelbefragung der Projektbearbeiter per Telefon oder mittels Fragebogen angestrebt, diese unterblieb aber, wenn die Internet-Dokumentation als ausreichend betrachtet wurde, wenn eine weitere Befragung keine besonderen neuen Erkenntnisse erwarten ließ und wenn bis zu einem bestimmten Stichtag keine Vereinbarung eines Termins oder der Befragungsmodalitäten gelang.⁵

Digitale Bestandserschließung als modulares Gesamtsystem

Innerhalb der untersuchten Projekte gab es hinsichtlich der durchführenden Institutionen, der Bestände, der Erschließungsansätze und der Zielstellungen so große Unterschiede, daß zu fragen ist, ob überhaupt Bestandsdigitalisierungen in Bibliotheken mit Findbucherstellungen in Archiven oder Datenbanken zu ethologischen Filmen in Forschungseinrichtungen verglichen werden können. Zusammengehalten wird die Untersuchungsmenge aber dadurch, daß alle Projekte darauf abzielen, große und in der Regel komplexe Informationsmengen unter Nutzung digitaler Techniken in neuer Form zugänglich zu machen. Verschieden sind im wesentlichen die Ausgangspunkte und grundsätzlichen Perspektiven, mit denen z.B. Bibliothekare, Archivare oder Wissenschaftler aus einzelnen Disziplinen an ihre Erschließungsprojekte herangehen und durch welche die Unternehmungen in ihrer Ausgestaltung dann auch geprägt sind. Es wird hier aber zu verdeutlichen versucht werden, daß die digitale Bestandserschließung mit Gewinn als ein konzeptionelles Gesamtsystem aufgefaßt werden kann. In einem solchen Gesamtsystem füllen unterschiedliche Projekte nur einzelne Bereiche unterschiedlich gewichtet aus, bieten jeweils eigene Lösungsansätze an und zeigen gleichzeitig auf, welche Bereiche in anderen Projekten noch offen sind oder nur anders gelöst werden. Im besten Fall kann es dazu führen, daß schon zu Beginn eines Projektes alle Arbeitsbereiche und möglichen Vorgehensweisen Eingang in die Gesamtkonzeption finden. Eine Reduzierung der Projektauswahl hätte folglich nur die Breite verschleiert, welche die Frage nach der digitalen Bestandserschließung eröffnet. Gleichzeitig wäre die Perspektive auf Probleme jenseits enger Insellösungen verstellt worden.

⁵ Die entstandene Datenbasis ist naturgemäß umfangreicher und detaillierter, als es hier wiedergegeben werden könnte. Der Autor steht deshalb für einzelne Nachfragen gerne zur Verfügung, z.B. per E-Mail unter Sahle@uni-koeln.de

Da die prinzipielle Offenheit für Erweiterungen und Veränderungen ein Grundcharakteristikum dieses Arbeitsfeldes ist, lassen sich alle Projekte durch ein Maximalmodell beschreiben. Die Überführung eines realen Bestandes in einen allgemein zugänglichen digitalen Bestand besteht grundsätzlich aus drei Komponenten: der wiedergebenden, der erschließenden (im engeren Sinne) und der öffentlich machenden. Anders ausgedrückt: ein realer Bestand kann zu einem erschlossenen digitalen Bestand werden, durch den Prozeß der digitalen Wiedergabe, den Prozeß der Sammlung von Erschließungsinformationen und den Prozeß der Publikation. Skizze 1 versucht diese Dreigliedrigkeit zu verdeutlichen.

1. Text	Gedruckt (hauptsächlich Bibliotheksmaterial)	Bücher	11
		Periodika	6
		Vermischtes und sonstiges	4
	Handschriftlich/ Typographisch (hauptsächlich Archivmaterial)	Vermischte Archivalienbestände	12
		Kodizes	3
		Fragmente	1
		Inschriften	1
		Briefe	2
		Akten und Aktenähnliches	4
		Nachlässe	3
	Exzerpte, Zettel, Belegstellen		6
	Noten		1
2. Bild, Realie	Einzelne Bilder	Gedruckt	4
		Fotografien	6
		Zeichnungen, Gemälde, Landkarten etc.	4
	Film		2
3. Ton	Tonbandaufnahmen		2
4. Vermischtes			1

Die Bereiche der Repräsentation und der Erschließung haben einen Aspekt der „Tiefe“. Auch bisher war bereits die Rede von „flacher Bestandserschließung“ und von „Tiefenerschließung“. Hörte die flache Erschließung in der Regel bei der Bestandsbeschreibung oder der Elementbeschreibung auf, so meinte Tiefenerschließung oft die Wiedergabe der Vorlage in regulierter Form (z.B. als Editionstext) und die Erschließung durch genauere Erläuterungen, Register und Zusatzinformationen. Tiefe bedeutet hier aber nicht nur ein mehr an Informationen (Vollständigkeit) und Zugriffsoptionen oder eine höhere Qualität der Wiedergabe (z.B. digitaler Abbildungen oder Texte), sondern auch, daß teilweise eines auf dem anderen aufbaut. Dies ist z.B. bei der automatischen Texterkennung der Fall, die normalerweise über das Einlesen digitalisierter Bilder geschieht. Auf der Seite der Erschließung im engeren Sinne sind eher unterschiedliche Erschließungsverfahren aufgelistet, die nicht unbedingt direkt aufeinander aufbauen. So haben Elementbeschreibung, Rekonstruktion der Binnengliederung und Verschlagwortung zunächst nicht unbedingt etwas miteinander zu tun. Im Gegensatz zu den Zusatz- oder

Sekundärinformationen werden sie aber dem realen Bestand entnommen, während die ergänzenden Informationen von außen kommen.

Bestandsrepräsentation kann die Erschließung (im engeren Sinne) erleichtern, teilweise auch automatisieren. Zusammenhänge bestehen hier vor allem zwischen der Erstellung elektronischer Texte und den daraus zu gewinnenden Registern oder Volltextindizes. Für die Elementbeschreibung und Verschlagwortung wäre ebenfalls an automatische Verfahren der Erkennung und Klassifikation zu denken. Da diese aber in der Praxis noch keine Rolle spielen, wurden sie hier auch nicht visualisiert.

Die einzelnen Bereiche des Modells werden nachfolgend näher beschrieben. Zunächst ist aber noch auf die Fragen einzugehen, von wem, warum und in welcher Form die gegenwärtigen Erschließungsvorhaben durchgeführt werden, und welche Arten von Beständen davon betroffen sind.

Organisation und Bestandsauswahl

Die Begründungen für die Durchführung von Projekten zur digitalen Bestanderschließung sind zahlreich, erweisen sich bei näherer Betrachtung aber als Variationen einer kleinen Zahl von Grundmotiven. Hier steht das Schlagwort vom „besseren Zugriff auf die Quellen“ im Vordergrund. Es umschließt eine ganze Reihe von Erwartungen, ausgehend vom Orts- und Zeitunabhängigen Zugriff, der nun direkter und gleichzeitig mehrfach möglich sei, über die mächtigeren Recherchemöglichkeiten bis hin zur leichteren Weiterverarbeitung. Grundlage dafür ist ein zweiter Komplex von Argumenten, der auf die rationaleren, tieferen, vielfältigeren und komplexeren Formen digitaler Aufbereitung, Repräsentation und Bereitstellung von Quellenmaterial verweist, angesichts derer die neuere Entwicklung als logische und konsequente Weiterentwicklung traditioneller Erschließungsformen erscheint. Wollte man die erste Argumentationsgruppe als „nutzerorientiert“, die zweite als „methodenorientiert“ (nämlich an den Grundideen der „Erschließung“) beschreiben, so wäre noch eine dritte hinzuzufügen, die „bestandsorientiert“ genannt werden könnte. Hier wäre der Gedanke des Bestandsschutzes ebenso zu nennen, wie der Wunsch nach Erschließung höchst komplexer Sammlungen, die den bisherigen Methoden nicht recht zugänglich waren. Verbunden damit ist das Ziel, auch diese einer breiteren und effizienteren Nutzung zuzuführen.

Neben den drei Gruppen „inhaltlicher“ Motive spielen auch institutionelle Gründe eine Rolle bei der Durchführung von Pilotprojekten im digitalen Bereich: Sie dienen hier u.a. als Machbarkeitsstudien für weitere Aktivitäten und dem Know-How-Erwerb. Vor allem besteht aber die Hoffnung, die Stellung der eigenen Institution durch innovative Leistungen, durch neue Fachkompetenz und durch originäre Angebote zu verbessern und gegen die Unwägbarkeiten des allgemeinen Einsparungsdruckes wie auch die zu erwartenden grundlegenden Veränderungen im Bereich bestandehaltender

Institutionen abzusichern. Gerade unter den größeren Bibliotheken setzt sich inzwischen die Auffassung durch, daß eine Beschränkung auf die traditionellen Arbeitsformen und Angebote mittelfristig mit einem massiven Bedeutungsverlust einhergehen würde.

Entsprechend stark sind sie mit Pilotprojekten im Bereich der digitalen Bestandserschließung vertreten. Hinsichtlich der Zuordnung nach Institutionstyp wurden aus der Untersuchungsmenge 30 Projekte von Bibliotheken, 29 von Forschungsinstituten (im weitesten Sinne) und 15 von Archiven durchgeführt, wobei sie gerade in den Bibliotheken oft im größeren Kontext einer allgemeinen Erweiterung und Umstellung der Arbeitsformen, Aufgabengebiete und Dienstangebote unter dem Paradigma der „neuen Medien“ stehen.

Für die Beschreibung und Analyse der Entwicklung der Methoden digitaler Bestandserschließung sind einige Beobachtungen hinsichtlich der Formierung und Organisation der gegenwärtigen Pilotprojekte von grundsätzlicher Bedeutung: Es handelt sich in der Regel um sehr umfangreiche, selektive, kooperative, mittelfristige, fremdfinanzierte und klar abgegrenzten Vorhaben.⁶

„Umfangreich“ bezieht sich dabei auf zwei Aspekte: die Breite der zu leistenden Erschließungsarbeiten und die Menge des zu erschließenden Materials. Die zusätzliche Ebene der digitalen Repräsentation, die tendenziell komplexere und tiefere Erschließung und die Tendenz zu interdisziplinärer und interinstitutioneller Kooperation führen zu einer Ausweitung der Arbeit. Anders formuliert: Die neuen Möglichkeiten lassen neue und weitergehende Anforderungen als selbstverständlich erscheinen, die in organisatorisch größer angelegten Vorhaben umgesetzt werden. Aber auch die inhaltliche Größe ist oft beeindruckend: Es werden in der Regel nicht kleine Bestände probenhalber oder nur teilweise erschlossen, sondern vollständige Sammlungen, z.B. von 40.000 orientalischen Handschriftenfragmenten, 100 laufenden Metern Archivmaterial, 150.000 Druckseiten Musiknotation, 250.000 Zeitschriftenseiten des 18./19. Jahrhunderts, 500 Stunden Tonbandaufnahmen, 290.000 Filmmetern, 2,5 Millionen Belegzetteln, 1,25 Millionen gedruckten Seiten zu einem bestimmten thematischen Bereich oder 1,3 Millionen Fotografien⁷. Die Gründe hierfür sind zum einen der Wunsch nach „realistischen“ Bedingungen für solche Pilotprojekte, auch was den Umfang betrifft, zum anderen der Gedanke, daß neue

⁶ Eine signifikante Minderheit von Projektträgern betreibt aber auch die digitale Erschließung nahezu ihrer gesamten (in der Regel begrenzten) Bestände, ohne weitere Kooperation, mit Eigenmitteln und in langfristiger Perspektive. Dies sind in der Regel Projekte von speziellen Forschungseinrichtungen, deren Arbeit zum großen Teil auf einen bestimmten Bestand ausgerichtet ist.

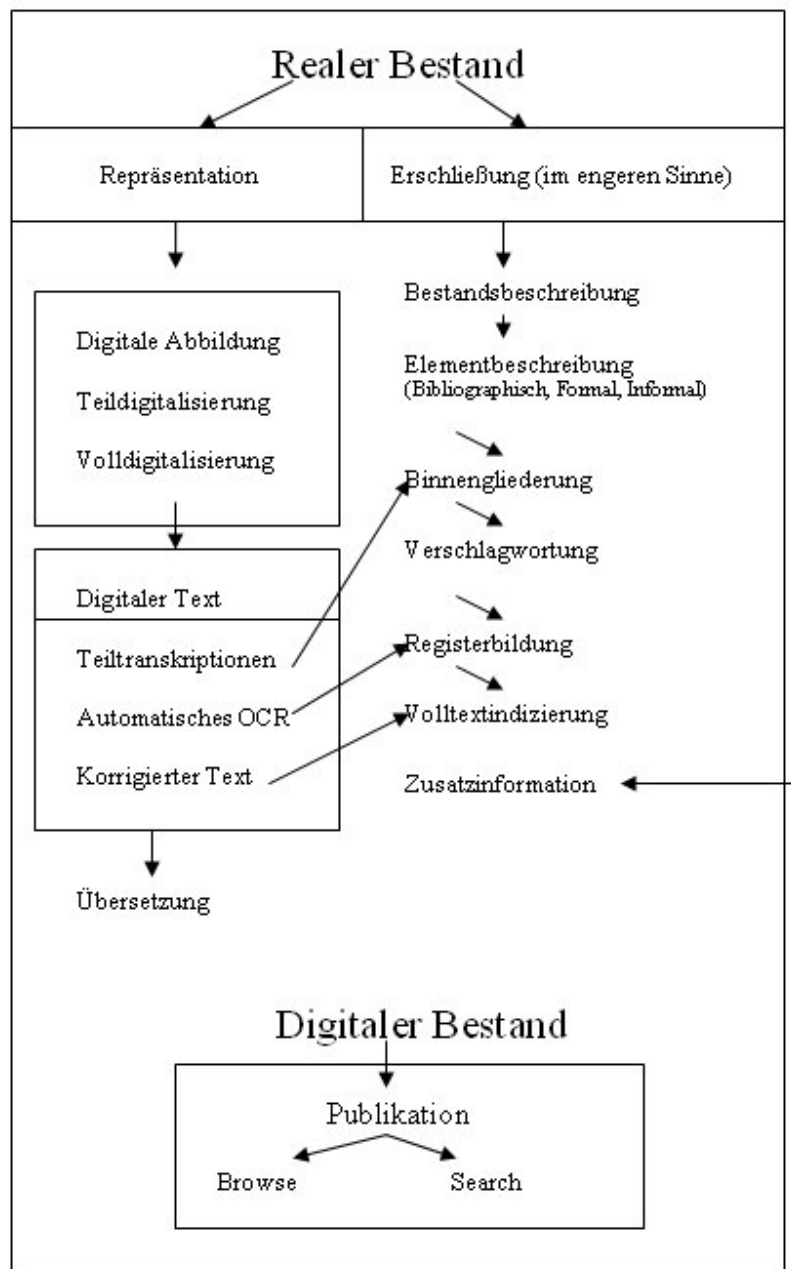
⁷ Die Beispiele ihrer Reihenfolge nach: „Turfanhandschriften“, „Der Archivbestand der Gesellschaft Naturforschender Freunde zu Berlin“, „Musikdrucke des 19. Jahrhunderts“, „Pädagogische Zeitschriften des 18./19. Jahrhunderts“, „Tonbandmitschnitte des Auschwitz-Prozesses“, „Humanethologisches Filmarchiv“, „Deutsches Rechtswörterbuch“, „Privat- und Prozeßrecht des 19. Jahrhunderts“, „Nationales Bildarchiv zur Kunst und Architektur in Deutschland“ vgl. jeweils die Internet-Dokumentation.

digitale Angebote eine „kritische Masse“ erreichen müßten, um von den Nutzern wahrgenommen zu werden und um bei ihnen eine Einarbeitung in die neuen Arbeits- und Benutzungsformen als sinnvolle Investition erscheinen zu lassen.

„Selektiv“ sind die Vorhaben, weil sie selten auf alle Bestände einer Institution zielen, sondern sich auf bestimmte, klar abgegrenzte Teile beziehen. „Kooperativ“ bedeutet, daß sie in Zusammenarbeit mit weiteren beständehaltenden Institutionen, Fachwissenschaftlern, Informatikern, externen Dienstleistungsunternehmen und (seltener) Verlagen durchgeführt werden. „Mittelfristig“ sind sie, weil sie als „fremdfinanzierte“, nämlich durch öffentliche Förderprogramme unterstützt, deren zeitlichen Rahmenvorgaben folgend zumeist eine Laufzeit von zwei bis fünf Jahren haben. „Klar abgegrenzt“ meint schließlich, daß sie zwar intern stark modularisiert, selbst aber nicht Teil eines größeren Projektes sind, sondern vielmehr ein eigenständiges und unabhängiges Informationsangebot erreichen sollen.

Für die Auswahl der zu erschließenden Sammlungen ist festzustellen, daß grundsätzlich „alle“ Arten von Dokumenten- oder Quellenbeständen Gegenstand von Erschließungsvorhaben sind. Das ist nur scheinbar trivial, in Wirklichkeit aber bemerkenswert, weil hierin ein Unterschied zu den traditionellen Erschließungsformen liegt. Bezogen sich jene vornehmlich auf eine bestimmte Gruppe von Beständen, die für ihre Methoden zugänglich waren und vernachlässigten sie dabei andere oder erschlossen diese nur sehr oberflächlich, so ist für die digitale Bestandserschließung fast von einer „Verobjektivierung der Quellenwahrnehmung“ zu sprechen. Dies betrifft nicht nur bildhafte Quellen oder Audiodokumente, die nun sogar leichter als textorientierte Quellen unmittelbar digital abzubilden sind, sondern auch sehr komplexe oder vermischte Bestände, die bislang weniger wahrgenommen und genutzt wurden, eben weil sie so schwer zu erschließen waren.

Andererseits könnte man für diese Phase der Entwicklung von einer Tendenz zu außergewöhnlichen, ja exotischen Beständen sprechen, wobei die Gründe auf der Hand liegen: Mit ihnen sind die Vorteile digitaler Arbeitsformen deutlicher zu zeigen, die notwendigen Drittmittel sind leichter einzuwerben und potentielle Nutzer sind hier eher von der Sinnhaftigkeit einer Umstellung der eigenen Arbeitsweise zu überzeugen eben weil es keine Alternative gibt.



Skizze 1: Gesamtmodell digitaler Bestandserschließung

Innerhalb der Untersuchungsmenge ergab sich folgende Verteilung (siehe Tabelle) auf verschiedene Arten von Beständen.⁸

Dabei sollten die scheinbar relativ hohen Zahlen im Bereich gedruckter Publikationen und diverser Archivalienbestände nicht darüber hinwegtäuschen, daß es sich auch hier in fast allen Fällen um Bestände handelt, die in irgend einer Weise „besonders“ sind bzw. ungewöhnlich hohe Anforderungen an den Erschließungsprozess stellen.

Bestandsrepräsentation

Die digitale Repräsentation zerfällt in der Regel in zwei verschiedene Formen der Wiedergabe. Hier ist eine unmittelbare digitale Abbildung (von Bild oder Ton) und die Erstellung eines elektronischen oder digitalen Textes zu unterscheiden. Dies gilt im Grunde für alle Bestandsarten, soweit sie Texte oder sprachliche Informationen im weitesten Sinne wiedergeben. Bei einigen anderen Arten gibt es allerdings keine sprachlichen Informationen,⁹ bei den anderen ist zu fragen, wie die Funktionsabgrenzung zwischen dem digitalen optischen Abbild bzw. dem elektronischen Text jeweils aussieht.¹⁰

Da die Digitalisierung von Audiomaterial sowohl selten als auch relativ einfach durchführbar ist, geht es im folgenden nur um die Konversion optischer Informationen. Die Grundfragen betreffen hier die Qualität (Auflösung und Farbtiefe), die Datenhaltung (Dateiformate und Speichermedien) und die Organisation (verwendete Technik und Outsourcing) einer Digitalisierungskampagne.

Entscheidungen für eine bestimmte Qualität der Wiedergabe werden einerseits von technischen und finanziellen Rahmenbedingungen bestimmt, andererseits von der Präferenz für eine von zwei grundsätzlich divergenten Zielstellungen geleitet. Hoch auflösende digitale Abbildungen sind teilweise nur mit sehr teurem Gerät zu erreichen und implizieren große Dateien, die Probleme bei der Speicherung aufwerfen, schwieriger zu bearbeiten sind und teil-

⁸ Hinsichtlich der Verteilung nach unterschiedlichen Geschichtszeiträumen ergaben sich folgende Zahlen: Ohne spezifischen Geschichtszeitraum oder umfassend: 2 Projekte, Antike 4, Mittelalter 11, 16./17. Jahrhundert 10, 18./19. Jahrhundert 26 und 20. Jahrhundert 18 Projekte. Wollte man hier hinsichtlich der tatsächlichen Überlieferung und des zu erwartenden Publikumsinteresse von einer Tendenz zu mittelalterlichen und frühneuzeitlichen/frühmodernen Beständen sprechen, so ließen sich als Gründe die gleichen anführen, wie für die Bestandsauswahl im allgemeinen, ergänzt noch durch Überlegungen zum Bestandsschutz.

⁹ Dies gilt vor allem für rein bildhafte Quellen wie Gemälde, gedruckte Illustrationen, Fotos oder Filme.

¹⁰ Für die optische Abbildung einer texttragenden Vorlage kann es drei Funktionen geben: (1.) Wiedergabe der bildhaften Informationen der Quelle (Layout, Struktur, Farben etc.), (2.) Ersatz für den Text, der nicht elektronisch vorliegt, aber über das Bild vom menschlichen Benutzer „gelesen“ werden können soll oder (3.) Vorstufe und Hilfsmittel für einen (automatisch) zu erzeugenden elektronischen/digitalen Text.

weise auch wegen der vermeintlich schlechten Datentransferzeiten bestimmter Publikationsmedien abgelehnt werden. Alle Gründe aus diesem Bereich sind von abnehmender Relevanz;¹¹ zusätzlich hat sich in der Praxis die Auffassung durchgesetzt, daß Entscheidungen über die Qualität in stärkerem Maße von inhaltlichen Überlegungen geprägt sein sollten, als von den jeweils aktuellen technischen Rahmenbedingungen.

Entscheidend ist letztlich die Frage, was digitale Abbildungen überhaupt leisten sollen. Hier gibt es zwei Grundpositionen, die zu ganz unterschiedlichen Anforderungen an die Qualität der Digitalisierung führen. Sehen die einen den Zweck in der Beigabe von wichtigen Informationen, wobei das Bild textliche und beschreibende Daten aus pragmatischen Gründen substituieren soll, so fordern die anderen, daß die elektronische Form der materiellen Vorlage so nah wie möglich kommen müsse, um ihr letztlich funktional gleichwertig zu sein und sie weitgehend ersetzen zu können. Die Berechtigung verschiedener Ansätze liegt auf der Hand, wenn man z.B. ein bibliographisches Periodikum der Mathematik mit mittelalterlichem Urkundenmaterial vergleicht. Ist der Informationswert im ersten Falle auf einen Text im engeren Sinne zu beschränken, so geht er im zweiten Falle offensichtlich darüber hinaus und umfaßt auch eine optische Ebene mit eigenem Informationsgehalt. Dies sollte aber nicht darüber hinwegtäuschen, daß es jenseits der klaren Fälle vom ersten Typ eher eine Frage unterschiedlicher Einstellungen zu Sinn und Zweck der Digitalisierung ist, als eine der verschiedenen Bestandsarten. Zur Illustration stelle man sich ein anderes Beispiel vor: Literatur des 18. Jahrhunderts ließe sich durchaus auf ihren „Text“ reduzieren, die bloße Lesbarkeit der Vorlagen in digitaler Konversion würde mithin den Zweck erfüllen. Es ist aber zu fragen, ob nicht die „Umgebung“ des Textes, nämlich der Einband und die Struktur des Buches, das Layout und die optische Oberfläche der Seiten, bis hin zum jeweiligen Zeichensatz und dem speziellen Aussehen des Papiers von Bedeutung ist, um das Dokument als umfassende und zeitgebundene Quelle wahrnehmen und seine Benutzung vor Ort ersetzen zu können.

Dies muß nicht das Ziel aller Digitalisierungsvorhaben sein. Eine Beschränkung auf den Transport der „wesentlichen“ Informationen einer Vorlage kann legitim sein und erlaubt geringere Auflösung und niedrigere Farbtiefe. Hier ist aber unbedingt vor einem Trugschluß zu warnen: Die Qualität der Digitisate darf nicht danach beurteilt werden, ob sie für das menschliche Auge „lesbar“ sind. Stehen sie zunächst als Ersatz für textliche und beschreibende Informationen, so sollte deren spätere maschinelle Erstellung nicht durch eine zu

¹¹ Die Preise für Scanner jeder Art fallen mit ihrer Verbreitung. Aufwendige Scanningkampagnen werden trotzdem immer kostenintensiv bleiben. Ein Verzicht auf die inhaltlich geforderte Mindestqualität führt langfristig aber nicht zu Einsparungen, da die Digitalisierungskampagne bei ausbleibender Benutzung zur Fehlinvestition würde oder teure Nachdigitalisierungen fällig wären. Die Probleme großer Dateien verringern sich angesichts der momentanen technischen Entwicklung (neue Speichermedien, leistungsfähigere Computer, bessere Netzleistungen) ebenfalls stetig.

schlechte Digitalisierung unmöglich gemacht werden. Grundsätzlich gilt, daß elektronische Texterkennung, jede Form von weiterer Verarbeitung oder andere Verwendungs- und Publikationsformen höhere Ansprüche an die Qualität der Digitalisierung stellen, als das lesende menschliche Auge vor einem Computerbildschirm. Aus diesen Gründen sollte selbst dann, wenn kein bis ins letzte gleichwertiges Abbild angestrebt wird, mit möglichst hoher Qualität gescannt werden.

Für die Farbtiefe gilt dies im gleichen Maße, wie für die Auflösung der Digitisate. Die Idee z.B. bitonal hergestellte Druckwerke auch bitonal zu scannen liegt nahe, greift aber zu kurz. Sie behindert eine spätere automatische Texterkennung und muß zusätzlich durch eine höhere Auflösung kompensiert werden. Hier besteht ein direkter Zusammenhang: eine höhere Farbtiefe transportiert weitere Informationen der Vorlage, die nicht nur einer Weiterverarbeitung zugute kommen, sondern auch eine höhere Auflösung ersetzen können.¹²

Umso schwerer sind konkrete Empfehlungen auszusprechen, mit welcher Auflösung und Farbtiefe denn nun gescannt werden sollte. Entlang der unterschiedlichen Zielvorstellungen und der Anforderungen der Dokumente kann nur für den Einzelfall entschieden werden. Das Gros der laufenden Projekte scannt mit 300 bis 400dpi, teilweise werden auch 600 dpi angestrebt, dies aber dann bei geringerer Farbtiefe.¹³ Zur Entscheidungsfindung sollte jeweils abgeschätzt werden, welche Anforderungen mittelfristig zu erwarten sein werden, sei es durch die Benutzer, sei es durch Verfahren der automatischen Weiterverarbeitung, wobei in beiden Bereichen auch verstärkt auf die Erfahrungen ähnlicher Projekte zurückgegriffen werden sollte.

Die prinzipielle Forderung nach möglichst guter Abbildungsqualität und nach Vernachlässigung aktueller technischer Hemmnisse macht es nicht unmöglich, schon jetzt Projekte so durchzuführen, daß von einer langfristigen Nutzbarkeit ausgegangen werden kann. Die technische Entwicklung und die wachsenden Ansprüche an die Qualität sind durchaus nicht uferlos und unkalkulierbar. Hier wird eine Grenze durch die begrenzte Schwierigkeit der Informationsgewinnung aus Dokumenten¹⁴ und durch biologischen Konstanten der menschlichen Benutzer¹⁵ vorgegeben, angesichts derer eine Entwicklung von

¹² Es ist deshalb in der Regel besser, eine scheinbar bitonale Vorlage mit 400dpi und 16 Graustufen zu scannen, als mit 600dpi und bitonal.

¹³ Eine nicht geringe Zahl von Projekten unterschreitet aber auch die Marke von 300dpi. Dies sind in der Regel schon länger laufende Verfahren, solche mit ausdrücklich geringem Qualitätsanspruch hinsichtlich der Abbildungen oder experimentelle Projekte, deren Schwerpunkt nicht auf dem Gebiet der Digitalisierung liegt.

¹⁴ Sofern es nicht extreme Spezialfälle von Vorlagen oder Auswertungspostulaten betrifft, ist ab einer bestimmten Grenze eine Erhöhung von Auflösung oder Farbtiefe nicht mehr mit einem automatischen Mehrgeinn an verwertbarer Information verbunden. Hier sind die Probleme dann eher durch intelligentere Software zu lösen.

¹⁵ Dies betrifft den stabilen Abstand zwischen Auge und zu betrachtendem Dokument/Monitor und die konstante Differenzierungsfähigkeit des Auges hinsichtlich einzelner Bildpunkte pro Flächeneinheit und hinsichtlich der Farbnuancen.

Auflösung bzw. Farbtiefe abzusehen ist, die gegen bestimmte Standardwerte konvergieren wird.

Im Bereich der Dateiformate und Speichermedien ist eine deutliche Entspannung hinsichtlich der scheinbaren Unwägbarkeiten der technischen Entwicklung bereits eingetreten. Mit TIFF hat sich ein nicht-proprietäres Industrieformat in der Praxis als Standard durchgesetzt, das nicht zuletzt durch seine hohe Verbreitung ein hinreichendes Maß an Zukunftssicherheit gewährleistet. Als Speichermedien einer Archivversion werden in der Regel CD-ROMS verwendet, deren Kosten kontinuierlich fallen. Ihre langfristige Haltbarkeit ist zwar noch umstritten, die Frage verliert aber an Relevanz, wenn in den entsprechenden Projekten bereits von Anfang an davon ausgegangen wird, daß in einem begrenzten zeitlichen Rahmen eine Migration auf die nächste Generation von Speichermedien vorgenommen werden wird und daß diese Generation im Verhältnis zu wachsenden Speicherkapazitäten relativ geringere Kosten verursachen wird.

Bei der technischen Durchführung von Digitalisierungskampagnen steht oft die Frage nach der geeigneten Scanner-Hardware und nach der bestmöglichen Organisation im Vordergrund. Im Hardwarebereich sollten einfache und kostengünstige Lösungen wie z.B. Flachbettscanner immer dann eingesetzt werden, wenn die Vorlagen es zulassen. Oft ist aber die mechanische Belastung, die mit diesem Scannertyp verbunden ist, nicht hinnehmbar und es kommen dann Spezialgeräte zum Einsatz. Hier sind zum einen digitale Kameras zu nennen, die in Verbindung mit Zusatzgeräten (z.B. spezielle Scantische, Fotolampen und Buchwippen) materialschonende und qualitätvolle Aufnahmen ermöglichen. Zum anderen wird oft der Umweg über die Mikroverfilmung genommen, die über Mikrofilmscanner zu den gewünschten Digitisaten führen. Mit beiden Typen entstehen heute bei hohen Qualitätsansprüchen noch sehr hohe Kosten, mit einem signifikanten Preisverfall ist aber für die nächste Zukunft zu rechnen.

Ein offener Punkt ist derzeit noch die Frage nach der Durchführung durch eigene oder zusätzliche Kräfte im Hause oder durch eine Auftragsvergabe an externe Dienstleister („Outsourcing“). In einer perfekten Welt müßte das Outsourcing geringere Kosten verursachen und zu besseren Ergebnissen führen, da Fremdfirmen über ein höheres Spezialwissen verfügen, die Geräte besser nutzen und eine effizientere Arbeitsorganisation gewährleisten könnten. In der Praxis zeigt sich aber unter den Bedingungen eines Arbeitsfeldes, das sich gerade erst entwickelt, daß es in vielen Fällen besser ist, die entsprechenden Arbeiten im Hause und unter eigener Kontrolle durchzuführen. Schlechte Erfahrungen mit der externen Auftragsvergabe waren in den bisherigen Pilotprojekten häufig zu beobachten und bezogen sich auf die Qualität, den Preis oder die Flexibilität der Dienstleistung, wie auch auf Reibungsverluste in der Kommunikation über die Anforderungen und auf Verzögerungen bei der Auftragsvergabe und Durchführung. Ein Mindestmaß an fachlichem Spezialwissen

im Scanningbereich muß von den Institutionen, die Erschließungsprojekte durchführen wollen, ohnehin erworben werden, um eine vernünftige Auftragsdefinition zu ermöglichen. Außerdem ist es gerade in einer Entwicklungsphase neuer Methoden wichtig, flexibel auf veränderte technische Möglichkeiten und neue Anforderungen reagieren zu können. Schließlich haben Pilotprojekte häufig einen so großen Umfang, daß in ihnen eine Anschaffung auch teurer technischer Geräte rentabel sein kann. Damit entfallen aber die wesentlichen Vorteile externer Dienstleister und es ist für den Augenblick als Zwischenfazit zu nennen, daß eine Auftragsvergabe nach außen dann sinnvoll ist, wenn die Anforderungen an Auflösung, Farbtiefe, Preis und Datenformat definitiv bekannt sind und ihre Erfüllung ebenso wie die Dauer der Durchführung gesichert ist. Das klingt selbstverständlich, ist es aber in der Praxis oft noch nicht. Gerade bei experimentellen Projekten, und als solche sind die meisten gegenwärtigen Pilotprojekte einzustufen, ist es oft mit negativen Begleiterscheinungen verbunden, wenn die Kontrolle über und Veränderungs- und Anpassungsmöglichkeiten im Scanningprozeß durch eine externe Auftragsvergabe aus der Hand gegeben wird.

Die optische digitale Abbildung von Dokumenten erhöht deren Zugänglichkeit und Benutzbarkeit und damit den „Wert“ eines Bestandes innerhalb einer bewahrenden Institution. Diese Argumentation, die im englischen Sprachraum unter dem Schlagwort des „add value“ geführt wird, erstreckt sich aber auch und vor allem auf die Herstellung elektronischer Texte als adäquate Repräsentation texttragender Dokumente in einer digitalen Arbeitsumgebung. Die optische Abbildung ist in dieser Hinsicht zunächst Substitut, dann aber auch Vorstufe für den elektronischen Text, der automatisch („OCR“) oder händisch („keyboarding“) gewonnen werden kann. In der Praxis werden nur in sehr vereinzelten Fällen größere elektronische Textmengen erstellt. Gerade das derzeit für Erschließungsprojekte herangezogene Material stellt in der Regel besonders hohe Ansprüche an eine OCR-Software oder menschliche Bearbeiter,¹⁶ so daß die Gewinnung eines fehlerfreien Textes nur mit äußerst hohem Kostenaufwand zu erreichen ist. Als Konsequenz wird leider meistens auf eine Texterstellung ganz verzichtet, oder sie wird auf einen unbestimmten späteren Zeitpunkt verschoben. Damit beginnt aber eine Schieflage zwischen den einzelnen Teilbereichen der digitalen Bestandserschließung, die auch für andere Felder zu beschreiben sein wird: Der Digitalisierung im Sinne optischer Abbildung wird eine sehr hohe Aufmerksamkeit zuteil, während man in anderen Bereichen noch auf bessere technische Lösungen wartet oder sich gar nicht erst

¹⁶ Wie bereits erwähnt, dominieren „ältere“, „komplexe“, „schwierige“ Arten von Beständen unter den laufenden Pilotprojekten. Gute Ergebnisse lassen sich bislang mit OCR-Programmen aber nur bei modernen Schriftsätzen und einem sauberen, nicht zu kleinem Druckbild erreichen. Die Lernfunktionen solcher Programme (ihre Trainierbarkeit) und die Fähigkeit, auch andere Schriftsätze - wie z.B. Fraktur - zu lesen, nimmt zwar zu, hat aber noch nicht den Punkt erreicht, an dem die notwendigen Kontrollen und Korrekturen die Finanzierungsgrenze der meisten Projekte unterschreiten würden.

mit gleicher Intensität um deren Entwicklung kümmert. Die digitale optische Abbildung bietet einen hohen Zugewinn an Benutzbarkeit und Zugänglichkeit. Der ebenso wichtige Schritt der eigentlichen Digitalisierung der textlichen Informationen der Dokumente sollte darüber aber auf gar keinen Fall außer Acht gelassen werden, zumal es hier auch um die Stimulierung und Entwicklung der Benutzungsformen geht, die nicht einseitig auf eine bildhafte Ebene beschränkt werden sollte.

Der hohe Aufwand zur Erstellung eines fehlerfreien elektronischen Textes sollte außerdem nicht den Blick darauf verstellen, daß bereits ein automatisch erzeugter, sehr fehlerhafter Text, zusätzliche Nutzungsmöglichkeiten bietet. Über „unscharfe Suchroutinen“ („fuzzy search“) kann er bei geringem Aufwand ein sehr leistungsfähiges Zugangsmittel zum Bestand sein, der eine Volltextsuche ermöglicht, ohne daß ein solcher Volltext im engeren Sinne überhaupt vorliegen würde.¹⁷ Als Ergebnis einer entsprechenden Anfrage an die Suchroutine wird denn oft auch eher die optische Abbildung des Dokuments angeboten, als der kryptische Zeichensteinbruch, den die automatische OCR erzeugt hat. Das Verhältnis von Aufwand und zusätzlichen Zugriffsmöglichkeiten ist aber so eindeutig, daß ein stärkerer Einsatz dieser Strategie nur allgemein zu empfehlen ist.

Was die technischen Strukturen und Dateiformate angeht, die für elektronische Texte verwendet werden, so steht die wünschenswerte Vereinheitlichung und Standardisierung noch am Anfang. PDF als Publikationsformat textlicher Daten ist zwar sehr einfach und komfortabel zu erstellen und zu benutzen. Hinsichtlich seiner Offenheit, Veränderbarkeit und Zukunftssicherheit und als proprietärer Standard ist es aber bedenklich. In der Praxis sollte es tatsächlich nur als Publikationsformat, nicht aber zur grundsätzlichen Organisation der Textdaten verwendet werden. Für diese sind als Grundformate ASCII (bzw. dessen Nachfolger als allgemeiner Standard UNICODE) zu empfehlen. Darüber hinaus ist auf die weitere Verbreitung von Auszeichnungssystemen zu hoffen, die für die einzelnen Bestandsarten gemeinsame Standards ermöglichen können.¹⁸

¹⁷ Siehe dazu vorbildhaft das Projekt der „Internet Library of Early Journals“ als Verbundprojekt mehrerer englischer Universitätsbibliotheken.

¹⁸ Für den Geisteswissenschaftlichen Bereich ist hier auf die „Text Encoding Initiative“ zu verweisen, die Richtlinien für die Auszeichnung verschiedener Dokumenttypen entwickelt. Leider wird darauf im deutschsprachigen Bereich noch relativ wenig zurückgegriffen. Außerdem existieren bislang nur Auszeichnungssysteme für philologische Quellen und Forschungszwecke. Ein drittes Problem betrifft die anstehende Umstellung bzw. Anpassung an die Entwicklungen, die mit der Etablierung von XML als neuem Leitstandard einhergehen.

Erschließung im engeren Sinne

Automatische Texterkennung ohne Fehlerbehandlung zur Ermöglichung weiterer Zugriffsmöglichkeiten dient weniger der Dokumentenwiedergabe, als ihrer Erschließung. Das gleiche gilt für partielle Textdigitalisierung. Hier werden in vielen Projekten nur Titel, Inhaltsverzeichnisse oder Register in eine elektronische Form gebracht, um die Dokumente auf diese Weise zu erschließen. Darüber hinaus kann auch eine Texterstellung unter Verwendung von Auszeichnungssprachen eine implizite Erschließung bedeuten, wenn die auf diese Weise dem Text hinzugefügten strukturellen und semantischen Informationen für den Zugang zum Material nutzbar gemacht werden.

Auch wenn die Bestandsrepräsentation so zur Erschließung beiträgt und die Übergänge fließend sind, ist an einer grundsätzlichen Unterscheidung festzuhalten. Erschließung „im engeren Sinne“ ist zu definieren, als regelgeleitete Sammlung von strukturierten Informationen zu einem Bestand, die zur Ermöglichung eines einfacheren und gezielten Zugriffs auf die enthaltenen Informationen beitragen sollen. Diese Informationen können zwar auch aus dem Bestand selbst gewonnen werden (wie oben angedeutet), folgen in der Regel aber einem äußeren Schema. Es handelt sich dabei um Daten, die Form, Inhalt und Struktur systematisch beschreiben, so daß man für die Erschließung im engeren Sinne auch die Umschreibung „Erhebung von Metadaten“ oder „Erhebung von beschreibenden Informationen“ benutzen könnte. Besonders der Begriff der Metadaten, als „Daten über Daten“, entspricht gut der Unterscheidung von Bestandsabbildung und Bestandserschließung (im engeren Sinne): Metadaten sind alle Daten, die nicht unmittelbar den Inhalt abbilden, sondern auf ihn verweisen. Sie müssen nicht (können aber sehr wohl) unmittelbar aus den inhaltlichen Daten gewonnen werden.¹⁹ Metadaten sind „erschließende“ Daten, weil sie durch Beschreibung oder Strukturierung die Verwaltung von Informationen und den gezielten selektierenden Zugriff auf sie ermöglichen.

Der Aspekt der Tiefe der Erschließung im engeren Sinne bezieht sich auf die Zerlegung des Bestandes in einzelne hierarchisch geordnete Einheiten und die Genauigkeit ihrer Beschreibung, die sich z.B. durch die Menge der erfaßten Merkmale oder Kriterien angeben läßt. Je nach Bestandsart kann das Vorgehen wieder sehr unterschiedlich sein und es werden jeweils andere Merkmale aufgenommen. Eine bloße Trennung der Erschließung nach inhaltlicher und formaler Erschließung entspräche zwar gewissen fachlichen Traditionen, wird aber weder der Komplexität der unterschiedlichen Erschließungsbereiche gerecht, noch würde sie interdisziplinäre Mißverständnisse verhindern. Auch in

¹⁹ Ein Beispiel dazu sind Wortkonkordanzen, die zwar aus den inhaltlichen Textdaten generiert, zum Zweck des gezielten Zugriffs aber regelgeleitet „umstrukturiert“ werden, indem sie z.B. in eine alphabetische Ordnung gebracht werden.

dem generellen Schaubild (Skizze 1) wird deshalb keine explizite Systematik der Erschließungsbereiche versucht, sondern es werden dort nur einige typische und gebräuchliche Erschließungsmittel aufgeführt.

Wollte man dennoch die unterschiedlichen Bereiche der beschreibenden Daten benennen, so könnten vier Gruppen gebildet werden:

- Abgrenzung und Nennung der Struktureinheiten: Dies kann von den einzelnen Teilen des Bestandes ausgehen (z.B. durch eine Beschreibung bibliographischer Einheiten, durch die Nennung von Aktentiteln, Beschriftung von Fotokartons etc.) und sich bei deren Untereinheiten fortsetzen (Aufsatz- oder Rezensionsteil einer Zeitschrift; Bescheid innerhalb einer Akte; Einzelne Aufsätze einer Zeitschrift, etc.)
- Formale und organisatorische Beschreibungen: Z.B. Angaben zum Material des Dokuments, Größe, Erhaltungszustand, Aufstellung/Signatur, Entstehung des Dokuments, Rechtsqualität, Erschließungsregeln, Bearbeiter etc.
- Inhaltliche Beschreibungen: Diese können entweder wieder auf ganze Dokumente oder Teile bezogen sein und normale Textform haben (z.B. als Regesten) oder begriffsorientiert und listenförmig sein (Register, Schlagwortlisten, Indizes). Hierbei sind jenseits der klassischen Felder „Person“, „Ort“, „Körperschaft“ etc. die unterschiedlichsten Kriterien denkbar, die sich aus der Besonderheit des Bestandes und den zu erwartenden Fragestellungen der Benutzer ergeben können.²⁰
- Anlagerung von Sekundärinformationen: Dies kann streng genommen auch schon bei der inhaltlichen Beschreibung geschehen, da diese oft durch Begriffe vorgenommen wird, die nicht unmittelbar der Vorlage entstammen. Hier ist aber weniger die begriffliche Klassifikation von Inhalten gemeint, als vielmehr die Hinzufügung ergänzender Informationen, z.B. biographischer Angaben zu den vorkommenden Personen, Verweise auf weitere Informationsmöglichkeiten etc.

Der Weg zu einer flachen oder tiefen Erschließung ist nicht durch diese Unterteilung vorgegeben, auch wenn bei einer flachen Erschließung am ehesten auf den dritten und vierten Bereich verzichtet wird. Minimalkonzepte der Erschließung können sich z.B. auf eine grobe Katalogisierung beschränken, die aber meistens auch schon einige formale Kriterien einschließt. Eine nächste Stufe könnte dann aber entweder die Strukturbeschreibung vertiefen oder aber

²⁰ Außerdem sollte das Kriterium auf die Breite des Bestandes anwendbar sein, um als sinnvolle Formalisierung angewandt werden zu können und eine Verbesserung des gezielten Zugriffs zu ermöglichen. Beispiele gibt es in fast jedem Erschließungsprojekt, sei es die Erfassung des Stechers bei Druckgraphiken des 16. Jahrhunderts, des Verfolgungsgrundes bei Entschädigungsakten oder der Variablen „Verhalten“ in humanethologischen Filmen.

zur Verschlagwortung übergehen. In Fällen tiefer Erschließung werden oft alle vier Bereiche abgedeckt und diese selbst wieder umfassend bearbeitet, indem z.B. eine ganze Reihe - eventuell auch hierarchisch gestaffelte - inhaltliche Kriterien aufgenommen werden. Die Tiefe der Erschließung läßt sich so anhand der vier Bereiche beschreiben: welche werden abgedeckt und wieviele Kriterien werden jeweils aufgenommen? Dabei nähern sich Erschließungsprojekte anderen wissenschaftlichen Formen wie Editionen und Auswertungen, die sich im Grunde nur durch eine noch größere Erschließungstiefe oder die Konzentration auf bestimmte Erschließungsperspektiven von flachen Bestandserschließungen unterscheiden.

Der Unterschied ist folglich ein gradueller und keiner, der die grundsätzliche Methodik oder die letztendlichen Ziele betreffen würde. Dies ist eine Eigentümlichkeit, die sich aus der Offenheit digitaler Erschließungsprojekte ergibt. Die meisten Vorhaben arbeiten zwar mit vorgegeben Erschließungstiefen, viele verzichten aber auch darauf, diese genau festzulegen und bei einigen gehört es geradezu zur Methode, sie offen zu lassen um dadurch auf die weitere Entwicklung der Technik, der Methoden oder der Benutzerwünsche reagieren zu können.

In der Praxis der untersuchten Projekte läßt sich eine große Bandbreite von Erschließungstiefen beobachten: von einfacher bibliographischer Erfassung bis hin zu extrem genauer und methodisch strukturierter Beschreibung z.B. von Periodika, Bildern oder Akten²¹ mit einer hohen Binnenstrukturierung, umfassender formaler, technischer und organisatorischer Beschreibung, einer Vielzahl erhobener Kriterien und Unterkriterien, der umfangreichen Anlagerung von Sekundärinformationen und teilweise auch der „Verlinkung“ mit externen Informationsressourcen im Internet.

Von der Durchführung her scheinen Bestandsrepräsentation und Metadatenerhebung umgekehrte Wege zu gehen. Wird bei der Repräsentation das Material in seiner ganzen Breite gleichmäßig in ein digitales Format konvertiert, aus dem dann ggf. zusätzliche Beschreibungen gewonnen werden, so setzt die Erschließung gezielt an einzelnen Punkten an, um Strukturen oder Inhalte zu systematisieren, die dann mit den Abbildungen verknüpft werden. Tatsächlich kann aber ein sehr enger Zusammenhang zwischen Repräsentation und formaler Beschreibung bestehen. So kann die Digitalisierung bereits unter Erschließungsaspekten durchgeführt werden, indem sie selektiv vorgeht und sich auf Elemente beschränkt, die für die Struktur des Bestandes wichtig sind (Titelblätter, Inhaltsverzeichnisse etc.), beim Digitalisierungsprozeß können be-

²¹ Vgl. als Beispiele für sehr ausdifferenzierte Tiefenerschließung die Projekte zu den „Dänischen Kulturzeitschriften 1917-45“, zum „Jahrbuch über die Fortschritte der Mathematik (1868-1943)“, zu den „DDR-Künstlerzeitschriften“, die REAL-Datenbank des Kremser Instituts für Realienkunde oder die beiden Projekte zu Entschädigungsakten in Münster und Hannover.

reits erste Metadaten mit erhoben werden²² und schließlich unterstützen digitale Abbildungen und Texte die Erschließungsarbeit, sowohl bei der händischen Eingabe von Daten, die teilweise „vom Bildschirm“ übernommen werden können, als auch bei automatisierten Verfahren.

Die Vergabe von Arbeitsaufträgen nach außen spielt im Bereich der Erschließung im engeren Sinne keine große Rolle. Fast durchgängig findet die Beschreibung der Dokumente durch hochqualifizierte Fachleute in der beständehaltenden Institution statt. Für die Organisation der Metadatenerhebung hat sich allerdings mit den digitalen Arbeitsformen und den neuen Kommunikationstechnologien die Option der stärker verteilten und interaktiven Erschließung ergeben. Diese Methode ist ein naheliegendes Mittel, mit dem Problem der schwierigen Organisation höchst umfangreicher Vorhaben umzugehen und nutzt die Offenheit der digitalen Erschließungsformen für eine permanente Veränderung und Erweiterung.²³

Von Standards bei der Metadatenerhebung kann in dreifacher Hinsicht gesprochen werden. Es können damit gemeint sein:

- Die Regeln der Beschreibung. Dies können die Regeln für Regestierung, alphabetische Ordnungen oder Schlagwortnormdateien sein.
- Die Strukturierung der erhobenen Daten, z.B. in relationalen oder hierarchischen Modellen, als Datenbankfelder oder Volltexte eventuell auch durch Auszeichnungssysteme.
- Die Formate der Datenhaltung, also Softwareformate, Grundformate, mögliche Austauschformate etc.

Alle drei Bereiche sind wichtig für die Frage der Zukunftssicherheit der geleisteten Erschließungsarbeit, für die Zusammenfassung mehrerer Bestände unter einer gemeinsamen Nutzeroberfläche und für einen rationellen Zugriff. Die offenkundig sinnvolle Entwicklung und Durchsetzung von Standards stößt in der Praxis aber auf einige Schwierigkeiten. So besteht ein Spannungsverhältnis zwischen dem allgemeinen Ziel der Verwendung von Standards und den speziellen Anforderungen des konkreten Bestandes, bzw.

²² Bei manchen Projekten werden z.B. bei der Herstellung von Bildern direkt Barcodes mit aufgenommen, die Angaben über das Dokument enthalten und später automatisch verarbeitet werden können.

²³ In der Praxis wird die Beschreibung oft von ganzen Gruppen vorgenommen. In einem Projekt gibt es sogar eine hoch organisierte Verteilung der Erschließungsarbeit auf eine große Zahl freiwilliger Experten, die ihren Beitrag über das Internet und E-Mail leisten (Vgl. das Projekt zum „Jahrbuch über die Fortschritte der Mathematik“). In anderen Projekten setzt man auf die Mitwirkung der Benutzer, die durch ihr Wissen zur korrekten und tiefen Beschreibung der Dokumente beitragen können. Hier kann ein entsprechendes Formular der Publikation des Bestandes direkt beigelegt werden, so daß Ergänzungen und Korrekturen entweder selbständig und automatisch, meistens jedoch nach einer Kontrolle durch Projektmitarbeiter, den Metadaten hinzugefügt werden können (Vgl. z.B. das „Bildarchiv der Deutschen Kolonialgesellschaft“).

wie diese vom jeweiligen Bearbeiter gesehen werden. Standards sind naturgemäß allgemeiner, als die konkreten Besonderheiten eines Bestandes, so daß sie oft nicht zufriedenstellend übertragbar sind. Die Neuentwicklung eines eigenen Regelwerkes oder einer eigenen Software scheint dann oft einfacher und letztlich präziser zu sein, als die Anpassung und Weiterentwicklung bestehender Systeme. Je tiefer die Erschließung, um so differenzierter und dem Bestand angepaßter wird das Erschließungssystem. Dies führt dann selbst bei sehr ähnlichen Arten von Dokumenten in ähnlichen Institutionen oft zu unterschiedlichen Vorgehensweisen. Daneben existieren bereits unterschiedliche Traditionen von Standards, die nun auch auf die digitalen Arbeitsformen einwirken. In Bibliotheken und Archiven haben sich jeweils andere Erschließungskonzepte entwickelt, die auch verschiedene Regelwerke und Systematiken zur Folge hatten. Dies hatte seinen guten Grund darin, daß beide Institutionstypen für unterschiedliche Bestandsarten zuständig waren. Inzwischen treten aber demgegenüber die gemeinsamen Schwierigkeiten und Zielstellungen stärker hervor, so daß die Chance, zu übergreifenden Methoden zu kommen, wächst. Gleiche Probleme im Bereich der Digitalisierung, der Datenorganisation und -verwaltung oder der Datenaufbereitung zur Publikation werden aber leider immer noch eher gemäß der erlernten Denkweisen und Grundstrategien angegangen, als zur Grundlage gemeinsamen Vorgehens gemacht.

Daß wir uns in einer Transmissionsphase der Arbeits- und Benutzungstechniken befinden führt grundsätzlich zu neuen Problemen, aber auch zu einer neuen Offenheit und damit neuen Chancen für umfassende Lösungen. Hier kollidieren Neuansätze, die allgemeine Systeme für beschreibende Daten entwickeln (vgl. RDF, Dublin Core), mit den Versuchen zur Weiterentwicklung und Übertragung etablierter Regeln und Muster aus den Bibliotheken und Archiven. Gegenwärtig erhöht sich dadurch noch die Zahl der Regelwerke und Systeme, anstatt sich der vereinheitlichenden Tendenz der Technik gemäß zu verringern

In der Praxis der untersuchten Projekte findet sich eine hohe Zahl unterschiedlichster Standards oder auch Beschreibungssysteme, die nur eine geringe Verbreitung haben oder sogar nur in einzelnen Projekten zur Anwendung kommen. Dies betrifft sowohl die Regeln für die Erfassung von Metadaten, als auch deren Struktur, als auch deren technische Formate. Leiten sich die unterschiedlichen Modelle zur Beschreibung von Dokumenten noch aus deren Besonderheiten ab, so ist die Divergenz der Datenorganisation schon weniger zwingend. Sie ist aber ähnlich hoch und selbst im Bereich der Dateiformate werden teilweise noch aufgrund der benutzten Software „Standards“ verwendet, bei denen nicht gesichert ist, daß in Zukunft ein Export der Daten in ein anderes System ohne Informationsverluste möglich sein wird. Das Bewußtsein für die Problematik ist zwar hoch entwickelt, bei der praktischen Durchführung wird aber

doch meistens anderen Gründen als dem Wunsch nach Einheitlichkeit bei der Entscheidung für oder gegen ein bestimmtes System der Vorzug gegeben.

Für die technische Umsetzung von Erschließungsprojekten wird in den meisten Fällen auf klassische Datenbankkonzepte zurückgegriffen. Das heißt, die beschreibenden Daten werden in einer relativ simplen Tabellenstruktur verwaltet, um Suchanfragen zu ermöglichen, an deren Ende die Verknüpfung mit den digitalen Abbildungen steht. Das alternative Modell inhärenter Erschließung mit Auszeichnungssystemen spielt dagegen derzeit noch eine geringere Rolle. Zu Anwendung kommt eine sehr hohe Zahl unterschiedlicher Software, von einfachen Datenbankprogrammen über eigens entwickelte Anpassungen, komplett neu entwickelten Werkzeugen bis hin zu umfangreichen und hoch spezialisierten Datenbank- und Dokumentenmanagementsystem. Die Entscheidung für oder gegen eine bestimmte Software wird dabei oft nicht durch deren möglichst exakte Eignung bestimmt, sondern von pragmatischen Gründen: Was ist bereits vorhanden? Wofür stehen erfahrene Anwender zur Verfügung? Motiv dafür ist, z.B. bei den Bibliotheken, oft auch der verständliche Wunsch nach Kompatibilität zwischen neuen Erschließungsprojekten und den bisherigen Formen elektronischer Kataloge, die möglichst alle von einer gemeinsamen Benutzeroberfläche aus zugreifbar sein sollen.

Publikation

Aber sind solche Schnittstellen als Publikationsform zu bezeichnen? Man könnte meinen, ein digitaler Katalog und ein digitales Findmittel - letztlich also nichts weiter als eine Suchmaske - sei keine Publikation. Wenn digitale Bestandserschließung aber als grundsätzlich offenes System aufgefaßt wird, dann ist ein digitaler Katalog die erste Stufe innerhalb einer umfassenden Publikation. Deshalb ist jede Form der Möglichkeit öffentlichen Zugriffs auf einen Bestand als Publikation zu betrachten und unter einer gemeinsamen Perspektive zu untersuchen. Von einem Aspekt der „Tiefe“ kann man in diesem Bereich kaum sprechen. In der Regel werden alle digitalen Informationen verfügbar gemacht, wenngleich die digitalen Abbildungen auch oft in schlechterer Qualität veröffentlicht werden, als sie eigentlich in der Archivversion vorliegen. Die verschiedenen Publikationsformen lassen sich eher danach unterscheiden, wie weit sie ihre Daten für einen möglichst einfachen Zugriff nochmals aufbereiten. Man könnte in diesem Sinne von „rudimentärer Publikation“ sprechen, wenn z.B. nur eine einfache Suchmaske mit einem Suchfeld als Benutzerschnittstelle angeboten wird oder von einer „umfassenden Publikation“, wenn verschiedene komplexe Suchroutinen ermöglicht werden, der Bestand auch auf anderen Wegen (browsing) benutzbar ist und durch zusätzliche Informationen und Hilfen dem Benutzer zugänglich gemacht wird. Dies betrifft

folglich nicht die repräsentierenden oder beschreibenden Daten selbst, sondern eher, was für die Veröffentlichung aus ihnen gemacht wird.

Denn die öffentliche Verfügbarkeit eines erschlossenen Bestandes ergibt sich nicht von selbst aus den vorhandenen digitalen Daten. Es ist nach der digitalen Abbildung und der Erhebung von Metadaten ein dritter Schritt nötig, der wieder spezielle technische, methodische und organisatorische Konzepte erfordert. Die beständehaltenden Institutionen, die meistens auch die Projektleitung haben sind zwar für die Publikation verantwortlich und stellen sie in der Regel zur Verfügung, sie erstellen sie aber nur, wenn sie mit einfacher Standardsoftware oder in Anknüpfung an bereits bestehende Publikationssysteme vorgenommen wird.²⁴ Jenseits dieser einfachsten Form der Publikation herrscht die Grundvorstellung der fachlichen Aufgabentrennung vor. Archivare, Bibliothekare und Fachwissenschaftler zählen die Frage nach der Form und der Erstellung der Bestandspublikation in der Regel nicht zu ihren Aufgaben und verfügen auch nicht über entsprechende Spezialkenntnisse und Erfahrungen.

Bei der Frage nach den Partnern in diesem Bereich wäre zunächst an die Verlage zu denken, die traditionell das Gebiet der Publikation, von der Methodik bis hin zur praktischen Erstellung und Verbreitung abdeckten. Für die digitale Bestandserschließung spielen sie aber kaum eine Rolle. Es besteht derzeit noch eine grundsätzliche Zurückhaltung der Verlage gegenüber digitalen Publikationsformen, für die auch bei ihnen kein Expertenwissen vorhanden ist, deren Marktchancen man nicht abschätzen kann und bei denen man deshalb vor ungewissen Investitionen zurückschreckt.²⁵ Die meisten Verlage warten noch ab, um die Entwicklung besser abschätzen zu können und publizieren - wenn überhaupt - mit möglichst geringem eigenen Aufwand eher schlechte Versuchsballons als umfassend ausgestaltete Multimedia-Produkte, wobei fraglich ist, ob so ein neuer Markt für digitale Publikationen im wissenschaftlichen oder Publikumsbereich entstehen kann.

Der Regelfall innerhalb der untersuchten Erschließungsprojekte ist eine Kooperation der beständehaltenden Institution mit EDV-Fachleuten aus verschiedenen Bereichen. Häufig sind dies Angehörige von Rechenzentren oder Informatik-Lehrstühlen, die den Institutionen ohnehin organisatorisch nahe stehen, wenn es sich z.B. um die Bibliothek und das Rechenzentrum

²⁴ Der typische Fall ist hier die Einfügung digital erschlossener Bestände in vorhandene elektronische Kataloge oder die Publikation aus umfassenden Dokumentenmanagementsystemen heraus, die eine Benutzerschnittstelle zur Verfügung stellen.

²⁵ Als Ausnahmen wären die Verlage Chadwyck-Healey und K.G.Saur zu nennen. Chadwyck Healey führt teilweise umfangreiche Projekte durch oder beteiligt sich an deren Finanzierung, um z.B. im Bereich der Bibliographien, der Literatur und Literaturgeschichte, der Quellensammlung oder Volltextdatenbanken inhaltsreiche und professionelle Publikationen - zu allerdings horrenden Preisen anzubieten. K.G.Saur übernimmt dagegen bereits digital erschlossenes Material aus den Projekten und vertreibt dies in relativ schlichter Form, gewissermaßen als Katalog-CD oder Bildsammlungen.

derselben Universität handelt. Die Frage der Publikation wird hier durch eine Kooperation gelöst, bei der die EDV-Spezialisten für die Entwicklung geeigneter Benutzerschnittstellen verantwortlich sind und die erschließenden Institutionen die Erstellung von Inhalten in der passenden technischen Form (z.B. in HTML, als Datenbankfile, etc.) übernehmen. Dies kommt zwar dem experimentellen Charakter der Pilotprojekte und ihrer Offenheit für Veränderungen entgegen, und auch die enge Verknüpfung von Bearbeitern aus dem konzeptionellen, dem erschließenden und dem technischen Bereich ist sicher positiv zu bewerten, es besteht aber die Tendenz, den Publikationsbereich nicht als eigenständige Aufgabe aufzufassen und hier keine eigenständigen Konzepte auszubilden, die über die Frage einfacher Zugänglichkeit hinausgehen.

Am Ende eines digitalen Erschließungsprojektes kann als eine der möglichen Publikationsarten ein Buch stehen. Warum auch nicht?! Bestände können digital erschlossen werden und die Ergebnisse dann in Druckform wiedergegeben werden. Einige Projekte im archivischen Bereich oder an Forschungseinrichtungen verfolgen dieses Ziel tatsächlich.²⁶ Es handelt sich aber um wenige Ausnahmen, denn der Informationsverlust, der wegen der unterschiedlichen Strukturen digitaler Medien und Druckmedien unweigerlich eintritt, wird in der Regel offensichtlich, sobald komplexere Erschließungsverfahren angewandt werden. Das gedruckte Buch oder eine andere Form von Ausdruck auf Papier ist als zusätzliches Derivat der digitalen Erschließung anzusehen. Für bestimmte Zwecke nützlich und ggf. leicht herstellbar, ist es aber der geleisteten Erschließungsarbeit insgesamt nicht angemessen.

Wenn hier von einem Derivat die Rede ist, so verweist das auf ein weiteres Kennzeichen digitaler Bestandserschließung. Sie hat die Tendenz zu multiplen Publikationsformen. Die Aufnahme von repräsentierenden Daten und ihre komplexe Erschließung erlauben eine Vielzahl unterschiedlicher Veröffentlichungsformen, in verschiedenen Medien, mit jeweils anderen Auswahlen und divergenten Benutzerschnittstellen.

Als „Standard“-Medium kann dabei derzeit das Internet, genauer das WWW, angesprochen werden. Fast alle untersuchten Projekte sind dort bereits einsehbar oder planen für die nächste Zeit eine Veröffentlichung. Dabei muß der allgemein und kostenfrei zugängliche Internetauftritt nicht immer das eigentliche Hauptziel sein. Teilweise dient er einfach der Projektdokumentation

²⁶ Ein naheliegendes Anwendungsgebiet für die parallele Veröffentlichung von Buch und Internet-Materialien sind Wörterbücher und Lexika. Diese werden oft weiterhin in gedruckter Form publiziert, zugleich wird aber die digitalisierte und mit komplexeren Beschreibungsverfahren erschlossenen Belegstellensammlungen im Internet zur Verfügung gestellt, die ganz anderen Nutzungsmöglichkeiten dient, als die gedruckte Variante. Siehe hierzu die Projekte zum „Deutschen Rechtswörterbuch“, zum „Chronologischen Wörterbuch des Deutschen“, die „Belegstellen zum Grimmschen Wörterbuch“ oder das „Altägyptische Wörterbuch“.

oder als Demonstrations- und Werbeversion für eine CD-Publikation oder einen kommerziellen Bereich im Internet. Für das WWW spricht, daß eine Publikation sehr einfach und schnell realisiert sowie leicht verändert und sukzessive ausgebaut werden kann. Die meisten Erschließungsprojekte arbeiten mit Datenbanken, die problemlos an das Internet anzubinden sind und für die eine Publikation sich auf die Konstruktion einer Benutzerschnittstelle beschränken kann, mit der jeder Interessent Zugriff auf die Daten erhält. Gleichzeitig kann das Internetangebot jederzeit verändert und ausgebaut werden, gemäß der weiteren Erschließung oder gemäß zusätzlicher Benutzungsmöglichkeiten durch ergänzende Datenbankschnittstellen oder zusätzliche Zugriffsarten wie strukturierte Darstellungen der Inhalte oder Sekundärinformationen.

Eine grundsätzliche Alternative zum Netz ist die CD-ROM. Teilweise als Träger anderer technischer Lösungen, zumeist aber wieder mit der gleichen technischen Grundlage als weiterentwickeltes oder spezialisiertes und auswählendes Derivat der Gesamterschließung. Die Vorteile der CD liegen dabei in ihrer Fähigkeit, relativ große Datenmengen mit einer Geschwindigkeit nutzen zu können, die das Internet derzeit nicht ermöglicht. Für die Publikation sehr hoch auflösender digitaler Bilder z.B. gibt es zur CD im Moment praktisch keine Alternative. Ein zweiter Vorteil ist, daß die CD-ROM besser in die etablierten Vertriebswege kommerzieller Güter paßt, da sie hier in dem Buch analogen Strukturen publiziert und vermarktet werden kann. Der noch ungeklärten Situation im Internet, bei dem man Bedenken hinsichtlich einer effizienten Kontrolle der Nutzung (verbunden mit der Frage nach Zugriff, Entgelt und Mißbrauch) hegt, steht ein weitgehend verlässlicher Rahmen der CD-Publikation gegenüber. Die CD ist allerdings nicht in gleichem Maße Probe- und Entwicklungsmedium sondern starrer Träger möglichst ausgereifter und umfassend gestalteter Publikationen, die z.B. in Zusammenarbeit mit Verlagen vertrieben werden könnten. Diese Kooperation ist in den meisten Erschließungsprojekten noch nicht gegeben und so ist auch die Erstellung von CD-ROMS zumeist eher eine langfristige Perspektive im Hintergrund der Projektkonzeptionen. Überhaupt ist die zukünftige Entwicklung hier noch offen, da unklar ist, wie die beständehaltenden Institutionen ihre Rolle im Publikationsbereich definieren werden. Neben einem verstärkten Engagement von Bibliotheken und Archiven bei der selbständigen Veröffentlichung ihrer digital erschlossenen Bestände, das dann auch voll entwickelte kommerzielle CD-ROMS umfassen könnte, stehen Abgrenzungsoptionen, die eine Aufgabenteilung mit den Verlagen, z.B. entlang der unterschiedlichen Medien (CD-ROM vs. Internet), der Kommerzialisierung (Grundversorgung vs. käufliche Weiterentwicklungen) oder der Tiefe der Erschließung (flacher Bestandszugang vs. umfassende Informationssammlungen) beinhalten könnten.

Wenn von multiplen Publikationen die Rede ist, so meint das nicht nur die verschiedenen Medien, sondern unter Umständen auch unterschiedliche Veröf-

fentlichungen im gleichen Medium. Dies kann z.B. abgestufte Nutzergruppen betreffen, die Trennung nach kommerziellem oder nicht kommerziellem Produkt und unterschiedliche Auswahlbereiche an Menge der Information oder Qualität der Wiedergabe.²⁷

Eine weitere Spielart der öffentlichen Informationsbereitstellung, bei der allerdings fraglich ist, ob man sie als Publikationsform bezeichnen sollte, ist die Abgabe von Daten oder Ausdrucken „on demand“, also auf den ausdrücklichen und gezielten Wunsch des Benutzers hin. Meistens handelt es sich dabei um eine Ergänzung, z.B. des Internet-Angebotes und umfaßt Formen, die eben nicht über das Netz zur Verfügung gestellt werden können oder sollen. Dies kann Bereiche betreffen, die nicht elektronisch erfaßt sind, also z.B. Kopien von Dokumenten, die zwar digital erschlossen, aber nicht digital abgebildet worden sind. Dies kann aber auch die hoch auflösende Variante von Abbildungen betreffen, die im Internet nur in geringer Auflösung angeboten werden. Dazu können z.B. CDs mit bestimmten Bildauswahlen auf Bestellung hergestellt und kostenpflichtig abgegeben werden.

Der Zugriff auf die Inhalte wird in der Regel durch Benutzerschnittstellen ermöglicht, die verschiedene Wege vorgeben, auf denen man zum erschlossenen Material vordringen kann. Hier sind zwei Grundtechniken zu unterscheiden. Zum einen die direkte Datenbankabfrage mittels Suchbegriffen („search“), zum anderen das Auswählen aus Listen oder das Verfolgen von Verknüpfungen („browse“). Das erstere könnte man als gezielte, planmäßige Suche beschreiben, die begriffs- und kategorienorientiert eine ganz bestimmte Auswahl aus dem Bestand produziert. Die Schnittstelle verhält sich bei einer Maskenanfrage interaktiv, weil sie in der Regel für jede Anfrage eine neue Antwort, z.B. in Form einer Trefferliste, generiert. Dagegen hat die Technik des „browsens“ eher heuristischen Charakter. Sie muß nicht zielgerichtet sein und es muß nicht bekannt sein, wie die Begriffe oder Kategorien eigentlich genau heißen, die gesucht werden. Die Anwendung verhält sich hier statisch, weil alle Informationen schon in einer festen Form, z.B. fest definierten Bildschirmseiten, vorliegen und der Benutzer zwischen diesen gewissermaßen nur hin- und herblättert.

Suchmasken und die mit ihnen möglichen Datenbankabfragen entsprechen sehr gut der Philosophie strukturierter Datenbanken, mit denen die meisten Erschließungsprojekte ihre beschreibenden Informationen verwalten. Meistens werden auf den unterschiedlichen Ebenen der Bestandsgliederung zu jeder Erschließungseinheit Metadaten erhoben, die einem bestimmten Kategorienschema folgen und oft auch einer festen Terminologie unterliegen. Dadurch entstehen gleichmäßig strukturierte Datensammlungen, die durch Programmroutinen leicht unter bestimmten Vorgaben durchsucht werden können. Such-

²⁷ Unterschiedliche Publikationsarten, mit verschiedenen Stufen der Zugriffsberechtigung und differenzierte Angebote je nach Nutzergruppen gibt es auch beim Projekt der ICE/FCR (Integrierte Computergestützte Edition/Fontes Civitatis Ratisponensis) in Graz.

masken erlauben es, alle Informationen einer Datenbank in beliebiger Selektion oder Verknüpfung von Kriterien und Begriffen auszuwählen und so eine benutzerspezifische Teilmenge herzustellen. Die Suche kann dabei nicht nur auf bestimmte inhaltliche Kategorien abzielen, sondern auch auf unterschiedliche Bereiche beschränkt werden. Hier wäre z.B. eine Suche im Volltext aller Dokumente von einer Suche in den Aufsatztiteln einer Zeitschrift zu unterscheiden.

Die Stärke von Suchmasken liegt in der schnellen Beantwortung höchst variabler Anfragen. Die Bearbeiter eines Erschließungsprojektes müssen nicht vorher wissen, welche Fragen einen Benutzer interessieren werden, sondern nur entscheiden, welche Kriterien am ehesten eine sinnvoll differenzierende Auswahl im Bestand ermöglichen, die den Fragestellungen des Benutzers entspricht. Unter einer Benutzeroberfläche können auch verschiedene Suchmasken angeboten werden. Oft bestehen neben den einfachen Routinen noch speziellere Masken, die eine Angabe von zusätzlichen Kategorien und Begriffen und dazu unterschiedliche Verknüpfungsarten zwischen ihnen (und, oder, und nicht, etc.) erlauben. Die Mächtigkeit der Suchmaschinen wird schließlich noch durch die Arbeit mit Platzhaltern gesteigert, so daß auch nach unterschiedlichen Schreibungen oder ganzen Wortfeldern gesucht werden kann.

Der Begriff „browse“²⁸ beschreibt dagegen die Benutzung einer nicht dynamischen Anwendung und ähnelt der Informationssuche, wie es sie auch bisher schon gab: das Durchstöbern von Katalogen, Listen oder Registern, das Blättern in Büchern oder Zeitschriften oder das Verfolgen von Fußnoten. Browsen wird zunächst ermöglicht durch hierarchisch strukturierte und geordnete Auflistungen des Materials. Auf diese Weise sind die gleichen Informationen darstellbar, wie sie in Datenbanken vorliegen. Solche Listen werden allerdings sehr schnell unübersichtlich und müssen einer Auswahl folgen, die sich auf ihre Ordnung und die Kombination von Kategorien bezieht. Sie ermöglichen einen oder mehrere Zugriffswege auf das Material. Dieser Zugriff folgt mehr oder weniger offensichtlichen Strukturen des Bestandes, ist dabei aber nur eine Auswahl aus den unzähligen möglichen Ordnungsstrukturen. So bleibt es dem Geschick der Bearbeiter überlassen, eine Systematisierung auszuwählen, die den Erwartungen möglichst vieler Benutzer entspricht. In der praktischen Anwendung wird der Zugriff durch fortschreitendes Auswählen in der hierarchisch jeweils nächsttieferen Ebene erreicht. Vom Zeitschriftentitel gelangt man z.B. zum Jahrgang, von diesem zu den Aufsatztiteln und von diesen zu den Aufsätzen selbst. Hier wird schon ersichtlich, daß z.B. für ein Browsen in einem festen Jahrgang (aber über mehrere Zeitschriftentitel) die Anlage einer

²⁸ Ich verwende den englischen Begriff als deutsches Lehnwort. Eine treffende deutsche Übersetzung scheint es nicht zu geben. Mögliche Übersetzungen wären: äsen, blättern, grasen, schmökern, überfliegen, weiden. Sinngemäß wäre ev. auch der Begriff „stöbern“, weil er einen Suchvorgang beschreibt, der nicht mit exakt definierten Zielen arbeitet.

neuen Listenstruktur erforderlich wäre. Einen browsenden Zugriff zu ermöglichen heißt immer, sich für die Anlage bestimmter Ordnungsmuster zu entscheiden.

Strukturierte Listen erlauben wie Datenbankabfragen einen Zugriff auf das gesamte erschlossene Material. Der Begriff des browsens erstreckt sich aber auch auf andere Techniken der Informationssuche. Die im Deutschen mögliche Übersetzung des Begriffes mit „stöbern“ beschreibt, daß auch das hin- und herblättern zwischen verschiedenen Dokumenten, das Suchen in thematischen oder anderen Auswahlen, das Lesen zusätzlicher Informationen, die Benutzung weiterer Hilfsmittel (z.B. von Lexika) und das Verfolgen von Verweisen und Hyperlinks Arten des Zugriffs auf das Material sein können.

In der Praxis laufender Projekte zur digitalen Bestandserschließung gibt es viele unterschiedliche Konzepte, wie der Benutzer zum erschlossenen Bestand geleitet werden soll. Entsprechend hoch ist die Zahl unterschiedlichster Benutzeroberflächen. Diese reichen von umfassenden Publikationen mit mehreren Suchmasken, Überblickslisten, thematisch geleiteten Hinführungen, Zusatzinformationen und anderen Angeboten bis hin zu einfachen Datenbankschnittstellen mit nur einem Eingabefeld für einen Suchbegriff. Tatsächlich ist die Benutzerschnittstelle und die Benutzerführung oft das schwächste Glied in der Kette der digitalen Bestandserschließung. Zwar sind fast immer alle erhobenen Daten auch in irgendeiner Weise erreichbar. Oft werden aber nur wenige Zugriffsmöglichkeiten angeboten und auf die Bereitstellung weiterer Auswahlhilfen oder zusätzlicher Informationen wird verzichtet, wobei die Auswahl nicht immer für jeden nachvollziehbar sein muß.

Eine Beschränkung auf nur eine der Benutzungsweisen „browsen“ oder „suchen“ ist oft zu beobachten, grundsätzlich aber abzulehnen. Beide Verfahren ermöglichen unterschiedliche Recherchestrategien. Auf eines zu verzichten bedeutet deshalb eine unnötige Reduktion der Benutzbarkeit eines oft mit hohem Aufwand erschlossenen Bestandes! Suchmaschinen sind mächtige Hilfsmittel bei der Recherche in großen Datenmengen und erlauben über die Auswahl und Kombination von Kategorien und Begriffen einen individuellen selektiven Zugriff, der über vorgefertigte Register und hierarchische Auswahllisten niemals ganz abgedeckt werden kann. Eine Beschränkung auf formalisierte Suchroutinen kann aber die Benutzung eines Bestandes auch erschweren! Treffer können nämlich nur dann erzielt werden, wenn die Vorstellungen von Kategorien und verwendeten Begriffen zwischen Benutzer und Bearbeiter übereinstimmen. Wer in einem Schlagwortregister nach „Bauer“ sucht, wird nichts finden, wenn die Erschließung mit dem Begriff „Landwirt“ gearbeitet hat. Das Problem hat verschiedene Dimensionen: Die Auswahl zwischen den Originalbegriffen der Quellen (in unterschiedlicher oder historischer Schreibweise) und einer externen Terminologie, zwischen festem Thesaurus und offenem je nach Bearbeiter und zwischen verschiedenen Sprachen, bei mehrsprachigen Beständen oder internationalen Projekten. Die verwendeten Schlagwortlisten und Thesauren of-

fenzulegen und möglichst direkt mit den Suchmasken zu verknüpfen, ist deshalb eine Mindestforderung, die alle Suchmasken erfüllen sollten. Erfreulicherweise sind gerade in letzter Zeit etliche Projekte zu dieser Praxis übergegangen. Neben den Eingabefeldern für Suchbegriffe können dazu die Listen mit den vorkommenden Begriffen aufgerufen werden. Unter diesen kann dann eine (auch eine mehrfache) Auswahl getroffen und direkt in die Anfrage übernommen werden.²⁹ Das Problem kann hier in der unüberschaubar großen Zahl der vorkommenden Begriffe liegen. Eine Lösung dazu wäre ihre Gruppierung oder hierarchische Ordnung.³⁰

Mit der Angabe von Begriffslisten kommt es zu einer Vermischung der Prinzipien des „browsens“ und „suchens“. Das Problem der Suchmasken besteht darin, daß nur etwas gesucht werden kann, von dem man bereits weiß - oder doch vermutet - daß es sich, und zwar genau in dieser Form, im Bestand finden wird. Dagegen unterstützt das browsen eine heuristische Form der Suche. Sie ermöglicht einen Überblick über das Material und läßt erkennen, was sich dort überhaupt finden lassen wird. Der Prozeß der Forschung besteht schließlich auch nur zur Hälfte aus der Beantwortung von Fragen. Die andere Hälfte ist die Entwicklung von Fragen und die Überprüfung, mit welchem Material sich die Fragen beantworten lassen werden.

Hinsichtlich der technischen Umgebung der Publikation spielen sogenannte „Autorensysteme“, mit denen z.B. ein großer Teil professioneller CD-ROMs gestaltet werden, im Bereich der digitalen Bestandserschließung keine Rolle, weil sie den Grundprinzipien von Offenheit, nicht-proprietärem Standard, leichter Veränderbarkeit und einfacher Anbindung an die zugrundeliegenden Datenbanken wie an das WWW als Publikationsmedium zuwiderlaufen. Es gibt statt dessen einen breiten Konsens über die Verwendung Internet-kompatibler Publikationsformaten, unter denen allerdings auch mehrere Varianten zur Verfügung stehen. Teilweise werden Publikationen im PDF-Format vorgenommen. Dieses erlaubt die schnelle Erstellung leicht benutzbarer konsistenter Publikationen, die auch gut ausgedruckt werden können. Die Möglichkeiten zur Gestaltung und Strukturierung komplexerer Informationsangebote sind allerdings beschränkt. Für tiefere Erschließungsformen und komplexere Informationsstrukturen ist PDF deshalb nicht geeignet. Der Regelfall digitaler Bestandserschließung sieht eine Verwaltung der repräsentierenden und beschreibenden Daten in unterschiedlichen Datenbanken oder mittels Auszeichnungssprachen (auf der Basis von SGML/XML/HTML) vor, deren öffentlicher Zugriff durch spezielle Benutzerschnittstellen ermöglicht wird.

²⁹ So beim „Bildarchiv der Deutschen Kolonialgesellschaft“ in Frankfurt und beim Projekt zu den „Annalen der Physik/Chemie“ in Jena.

³⁰ Das „Bildarchiv der Deutschen Kolonialgesellschaft“ in Frankfurt arbeitet z.B. mit einem solchen hierarchischen Thesaurus für geographische Begriffe.

Das Grundprinzip ist immer das gleiche: Man kann unterscheiden zwischen den grundlegenden repräsentierenden und erschließenden Daten und ihrer Publikation. Die ersteren liegen in den unterschiedlichsten Formaten und Strukturen vor, werden in einem arbeitsaufwendigen Erschließungsverfahren erstellt, können grundsätzlich verändert und ergänzt werden, bleiben aber tendenziell stabil. Die Publikation ist dagegen variabler: Es können verschiedene Publikationen parallel zur Verfügung gestellt oder sukzessive ergänzt oder gegen neue ausgetauscht werden. Dies entspricht gut den unterschiedlichen Entwicklungsständen von Digitalisierungs- bzw. Erschließungsverfahren auf der einen Seite und den digitalen Publikationsformen auf der anderen Seite. Eine Entwicklung der Formen der Veröffentlichung ist so unabhängig von den einmal festgelegten Strukturen der Digitalisierung und Erschließung möglich. Dazu ist es aber auch nötig, das Verständnis von „Publikation“ nicht auf eine einfache Suchroutine zu reduzieren, die nur beschränkte Zugriffswege öffnet. Publikation ist vielmehr als umfassendes Informationsangebot aufzufassen, das einerseits das vorhandene Material (repräsentierende und erschließende Daten) in möglichst guter Qualität und vollständig wiederzugeben in der Lage ist, und dies andererseits gemäß den unterschiedlichen Erwartungen der Benutzer tut. Dies schließt die Bereitstellung multipler Zugriffswege (Übersichten, Auswahllisten, einfache Suchmasken, komplexere Suchmasken) ebenso ein, wie begleitende Erläuterungen und Hilfestellungen. Ein hohes Maß an Automatisierung in der Veröffentlichung (wie es z.B. in den „Komplettlösungen“ der Dokumentenmanagementsysteme angestrebt wird) erhöht die Effizienz und die Kompatibilität von Erschließungsprojekten, ist aber der Weiterentwicklung nutzergerechter Publikationsformen nicht unbedingt zuträglich, da die Veröffentlichungsform eng an die Struktur der Datenhaltung angebunden bleibt, während eine umfassende Publikation eben mehr bieten sollte, als eine einfache Wiedergabe der Daten über wenige starre Kanäle.

Für die Publikation visueller Informationen werden die Archivversionen der Digitisate zumeist in eine niedrigere Auflösung und andere Dateiformate konvertiert. Zum einen bietet eine Reduktion der Qualität einen impliziten Schutz vor Mißbrauch, zum anderen kommt man so der beschränkten Übertragungskapazität des Internets entgegen. Bilder werden außerdem oft in verschiedenen Auflösungen angeboten, um z.B. über „thumbnails“ einen ersten schnellen Überblick zu ermöglichen und je nach Bedarf weitere Darstellungen in unterschiedlicher Dateigröße anzubieten. Als Format wird in der Regel JPEG gewählt, manchmal auch GIF, die neuere Alternative PNG findet noch fast keine Anwendung.

Zur Refinanzierung von Erschließungsprojekten über ihre Publikation sind eine ganze Reihe von verschiedenen Modellen denkbar. Im Internet könnten z.B. sehr wohl Gebühren für den Aufruf einzelner Angebote erhoben werden. Ein „pay per view“ entspricht allerdings noch nicht dem technischen Entwicklungsstand. Einfacher zu realisieren ist derzeit noch ein Abonnementsystem,

bei dem der Kunde für einen allgemeinen Zugriff zahlt und ihm dieser durch ein spezielles Paßwort ermöglicht wird. Noch einfacher ist der Vertrieb kommerzieller CD-ROMS, z.B. über die etablierten Wege des Buchhandels. Hierbei kann es auch eine Verknüpfung mit Internetangeboten geben. Mit dem Erwerb einer CD kann z.B. ein Zugriff auf Paßwort-geschützte WWW-Bereiche verbunden sein, die auch update-Funktionen wahrnehmen können. Ein Beitrag zur Refinanzierung kann ebenfalls durch die Abgabe von speziellen CDs oder Ausdrucken auf Papier - je nach den besonderen Wünschen eines Benutzers - geleistet werden. Dies entspräche den traditionellen Dienstleistung und könnte ohne weiteres in die Gebührenordnung einer Bibliothek oder eines Archivs eingefügt werden. Schließlich bietet die Aufgabenteilung z.B. mit einem Verlag eine weitere Möglichkeit, die für viele Projekte zumindest in der mittel- und langfristigen Planung eine Rolle spielt: Die beständehaltende Institution nimmt dabei die Digitalisierung und eine erste flache Erschließung vor und übergibt einem Verlag diese Vorarbeiten zu einem bestimmten Bereich zur weiteren Ausarbeitung und kommerziellen Verwertung, an deren Ertrag die Institution dann wieder zu beteiligen ist.

Dies entspricht der Grundtendenz, aus einem digital erschlossenen Bestand verschiedene Publikationen zu entwickeln. Die oben genannten kommerziellen Modelle schließen sich so wenig gegenseitig aus, wie parallele kostenfreie Angebote. Oft wird es wohl zu einem Szenario kommen, bei dem ein flaches Internet-Angebot mit niedrigen Bildqualitäten unbeschränkt zugänglich sein wird, während gleichzeitig vertiefende Publikationen zu ausgewählten Teilen des Bestandes auf CD-ROM vertrieben und ein umfassend ausgearbeitetes und qualitativ hochwertiges Informationssystem im WWW nur gegen Entgelt genutzt werden kann.

Im Vergleich zur Digitalisierung und zur Erschließung im engeren Sinne ist die Publikation der Ergebnisse insgesamt der am wenigsten entwickelte Bereich der digitalen Bestandserschließung. In der Regel liegt hier nicht der Schwerpunkt eines Projektes bzw. wird die Art und Weise, in der ein erschlossener Bestand schließlich öffentlich zugänglich gemacht wird, nicht als integraler Bestandteil der Erschließung aufgefaßt.³¹ Die Konzeption von Erschließungsprojekten und die Leitung der Durchführung liegt in der Regel bei der beständehaltenden Institution, die sich oft für die konzeptionelle Ausgestaltung der Publikation nicht zuständig fühlt. Gleichzeitig wird diese Aufgabe

³¹ Hier gibt es aber auch eine ganze Reihe von Gegenbeispielen. Siehe z.B. die Seite der Warburg Electronic Library (<http://www.sts.tu-harburg.de/projects/WEL/ublickfed.html>) zu diesem Thema. Weitere Projekte mit großem Gewicht auf der Publikation und Vermittlung auch das „Electronic Archive of Early American Fiction“, „CAMENA“, die „Flugschriften der Revolution von 1848“ oder die „Gutachten der Film-Oberprüfstelle“ (vgl. jeweils die Internet-Dokumentation zu diesem Artikel). Ganz besonders interessant ist in diesem Zusammenhang die Kremser REAL-Datenbank, die ein Beispiel dafür ist, welche vielfältigen und komfortablen Zugriffswege mit ein wenig Phantasie aus einer Datenbank generiert werden können!

aber auch nicht von anderen Projektpartnern abgedeckt. Die beteiligten EDV-Fachleute würden zwar die technische Umsetzung besorgen, sind aber keine Fachleute für die Entwicklung von Vermittlungskonzepten. Die Zurückhaltung der Verlage bremsen einerseits die Entwicklung multimedialer Publikationsformen, hat aber andererseits auch positive Nebeneffekte: Da ihre Tendenz zu Erzeugnissen, die möglichst einfach herzustellen und publikumswirksam sind, hier kaum eine Rolle spielt, bliebe für den wissenschaftlichen Ansatz innerhalb der Erschließung genug Raum, um ein gewisses inhaltliches Mindestniveau und die Verwendung von geeigneten Standards zu etablieren.

Ein wichtiger Grund für die geschilderte Situation ist das Fehlen ausgebildeter und etablierter Publikationsformen, an denen man sich orientieren könnte. „Das Medium hat noch nicht zu seiner Form gefunden“ hört man zuweilen als reichlich metaphysische Beschreibung, die aber zu Recht auf das Dilemma hinweist, daß allgemein verbreitete „Standardformen“ noch fehlen. Zugleich fehlt es an vielen Stellen aber auch an der Bereitschaft, zu deren Entwicklung beizutragen. Der Bereich der digitalen Bestandserschließung gehört zweifellos dazu.

Fazit

Wer heute von Bestandserschließung redet, meint damit etwas anderes als noch vor einigen Jahren. Durch den Wandel der Möglichkeiten impliziert der Begriff heute anderes und mehr als früher. War bis vor einiger Zeit damit die Erstellung eines Kataloges oder eines Findbuches gemeint, so assoziiert man heute tiefer gehende Methoden und Zielvorstellungen. Die digitale Bestandserschließung ist ein umfassendes Arbeitsgebiet mit neuen Akteuren, einer neuen Rollenverteilung, neuen Gegenständen und neuen Zielen geworden. Sie hat eine Tendenz zu größerer Tiefe entwickelt, zur digitalen Repräsentation des Bestandes in Ergänzung der beschreibenden Daten und schließlich auch zur eigenständigen - dabei teilweise multiplen - Publikation von repräsentierenden und beschreibenden Daten zu einem Bestand.

Damit ist die Frage der Abgrenzung neu gestellt, z.B. gegenüber der „Edition“ von Dokumenten, die Teil eines Bestandes sind. Auch Edition bedeutet nichts anderes als wiedergebende Erschließung, und so zeigt sich, daß Bestandserschließung und Edition unter einer ähnlichen Perspektive zu untersuchen sind und sich im Grunde vor allem im Umfang der Dokumentenmenge und in der Tiefe der Erschließung unterscheiden. Edition ist umfangsbeschränkte und tiefe Erschließung. Edition ist aber auch die logische Fortsetzung offener Erschließung, und so ist nicht immer klar anzugeben, wo die Erschließung aufhört und die Edition beginnt.³²

³² Ein Beispiel hierfür sind die Fontes Civitatis Ratisponensis innerhalb des ICE-Projektes in Graz, die eher als Quellenedition aufzufassen sind, denn als flache Quellenerschließung.

Aufgrund der prinzipiellen Offenheit der Erschließung innerhalb eines Maximalrahmens gewinnen Projekte in diesem Bereich außerdem immer stärker modularen Charakter. Sie sind Teilarbeiten einer Gesamterschließung - unabhängig davon, ob diese jemals erreicht wird oder überhaupt erreicht werden soll.

In der Konzeption der untersuchten Erschließungsprojekte gibt es so fundamentale Unterschiede, daß zu fragen ist, ob es überhaupt Sinn macht, sie unter einer vereinheitlichenden Perspektive betrachten zu wollen, oder ob man nicht besser die unterschiedlichen Vorgehensweisen jeweils nach den einzelnen Institutionstypen, Bestandsarten oder Grundzielen analysiert. Ob man also nicht besser weiterhin archivische Erschließung von bibliothekarischer Erschließung, die Bearbeitung von gedruckten Büchern von Bildern oder Tonaufnahmen, die Findmittlerstellung von einer umfassenden Multimediapublikation trennt. Genau das wäre aber angesichts der zunehmenden Ähnlichkeiten in einem neuen, erweiterten und grundsätzlich offenen Arbeitsbereich fatal. Die speziellen Probleme und Lösungsansätze, z.B. der Gestaltung von Metadatenrastern, sind nur noch relativ wenig maßgebend gegenüber gemeinsamen Fragen, z.B. nach Standards, der langfristigen Sicherung oder den unterschiedlichen Publikationsformen. Bibliotheken und Archive haben bislang teilweise für ähnliche Bestände unterschiedliche Beschreibungsverfahren entwickelt. Spätestens bei technischen Fragen, wie der Bildauflösung, verschwinden solche Unterschiede aber und es zeigt sich ein zunehmendes Maß an gemeinsamen Problemen.

Für die Praxis der vorhandenen Projekte ist nicht zu leugnen, daß deren Konzeptionen nicht von einer fachübergreifenden informationstheoretisch orientierten Sichtweise ausgehen, sondern von verschiedenen fachspezifischen Ausgangspunkten, wodurch sie unterschiedlichen Traditionen verhaftet sind. Hinter etlichen Projekten steht nicht so sehr der Gedanke, auf einem ganz neuen Feld etwas grundsätzlich neues erfinden zu müssen, sondern die erlernten Strategien und die bekannten Ziele nur sukzessive um neue Möglichkeiten zu erweitern. Digitale Bestandserschließung kann auch mit einer Zielvorstellung durchgeführt werden, die sie als erweiterten Katalog, als erweitertes Findmittel, als umfangreiches Bilderbuch, als vertieftes Informationssystem, als multimediales Buch erscheinen läßt. Die konzeptionelle Entwicklung geht de facto von etablierten Mustern aus und erweitert diese nur um naheliegende technischen Zusatzmöglichkeiten. Dies muß aber weder schlecht sein - den Benutzern geht es auch nicht anders - noch sollte man die Bandbreite höchst unterschiedlicher Zielsetzungen unterschätzen, bei denen es so fundamentale Alternativen gibt, daß sie zwangsläufig zu anderen Konzeptionen führen müssen.

Diese höchst unterschiedlichen Perspektiven werden durch den zunehmenden Umfang der Projekte und die höhere Zahl der beteiligten Fachleute noch erweitert. Findet eine Kooperation mit einem Verlag statt, so fließt auch dessen Sichtweise in die Konzeptbildung ein. Werden Aufgaben an externe Dienstlei-

ster vergeben, so können auch hier andere Vorstellungen die Projektgestaltung beeinflussen. Die Konfrontation divergenter Perspektiven wird bei der Zusammenarbeit mit Fachleuten aus dem EDV-Bereich besonders deutlich. Auch diese bringen eigene Vorstellungen und Konzepte mit, die sich teilweise fundamental von denen unterscheiden, die auf der „inhaltlichen“ Seite, also bei Bibliothekaren, Archivaren, Dokumentaren oder anderen Fachwissenschaftlern herrschen. Im Grunde haben es alle mit dem Problem zu tun, wie Informationen zu erheben, zu strukturieren, zu verwalten und öffentlich zugreifbar zu machen sind. Was sie im einzelnen aber darunter verstehen und wie es von ihnen konzeptionell begriffen wird, kann so unterschiedlich sein, daß selbst ein kommunikativer Austausch darüber schwierig ist. Ein Denken in so unterschiedlichen Kategorien, wie „Datensatz“ versus „Quelle“, „Suchroutine“ versus „Fragestellung“, „Datenbank“ versus „Textkorpus“ oder „Variablendefinition“ versus „unregelmäßige Orthographie“ führt zwangsläufig zu Verständigungsschwierigkeiten, wiewohl es auch gegenseitig befruchtend sein kann. Das Problem besteht darin, daß unterschiedliche Gruppen von Spezialisten es mit einem gemeinsamen Problem zu tun haben, hinsichtlich der Natur des Problems aber eine unterschiedliche Perspektive mitbringen, die sich aus ihren unterschiedlichen Erfahrungen und Denkmodellen ergibt. In der Informatik wird dies manchmal als „Impedanz-Fehler“³³ bezeichnet, der nicht nur zwischen Geisteswissenschaftlern und Informatikern, sondern zwischen jeder Gruppe hochspezialisierter Fachleute auftritt, und das Phänomen beschreibt, daß sich die Vorstellungen oder Anforderungen aus einem wissenschaftlichen System nicht ohne weiteres eins zu eins in ein anderes übertragen lassen.

Nicht nur in den konkreten Arbeitsschritten, sondern auch in der Konzeptbildung findet - vielleicht auch wegen des fehlenden gemeinsamen Rahmens - eine Modularisierung statt. Unter Verzicht auf die Entwicklung einer umfassenden Zielvorstellung werden einzelne Bausteine gebildet, deren Ausarbeitung voneinander getrennt und im wesentlichen den dafür vermeintlich kompetenten Spezialisten überlassen wird. Ein häufig zu beobachtender Fall ist die Konstruktion der Benutzerschnittstellen, die teilweise so vollständig informationstechnischen Fachleuten überlassen wird, daß sich die eigentlichen inhaltlichen Anforderungen kaum noch darin verwirklicht finden. Ein anderer typischer Fall ist die konzeptionelle Trennung von Digitalisierung, Beschreibung und Publikation. Es gibt derzeit eine ganze Reihe von Projekten, die zunächst digitalisieren, ohne bereits ein Konzept für die beschreibende Erschließung fertig ausgearbeitet zu haben. Noch häufiger steht das Beschreibungssystem fest, ohne daß geklärt wäre, in welcher Form die Daten überhaupt öffentlich zugänglich gemacht werden sollen.

³³ Siehe z.B. Grady Booch: Objektorientierte Analyse und Design. 3. Aufl., Bonn 1997, S. 20.

Modularisierung ist ein wichtiges Prinzip in einem Arbeitsgebiet, in dem die jeweiligen Vorhaben die Arbeitskraft und vielleicht auch die konzeptionellen Möglichkeiten der einzelnen Akteure übersteigen. Es wäre aber fatal, wenn sich Projekte auf allzu kleine Aspekte beschränken würden und wegen einer fehlenden Gesamtperspektive (für das Projekt) schließlich nicht weiter entwickelt werden könnten und damit ihre langfristige Benutzbarkeit gefährden würden. Ohne ein grobes Gesamtkonzept ist keine sinnvolle Modularisierung möglich! Ohne zu wissen, was eigentlich das Ziel eines Erschließungsvorhaben ist, kann keine vernünftige Entscheidung für die Durchführung bestimmter Teilbereiche getroffen werden. Ohne den Gedanken an die Benutzung und Publikation läuft die Gestaltung der Metadatenerhebung und ihre technische Organisation Gefahr, nur einen Datenfriedhof zu produzieren.

Ohne den Gedanken an die weitere Verwendung, Verarbeitung oder Veröffentlichung kann sich eine Digitalisierungskampagne als Fehlinvestition erweisen. Das Spannungsverhältnis zwischen der Offenheit für weitere Entwicklungen und Veränderungen auf der einen Seite und einem notwendigen groben Gesamtkonzept auf der anderen Seite wird sich nicht auflösen lassen. Es ist nur davor zu warnen, eine der beiden Seiten allzu absolut zu setzen. Offenheit und Veränderbarkeit sind vielleicht die beiden größten Stärken digitaler Arbeits- und Publikationsformen. Sie dürfen aber nicht mit Orientierungslosigkeit verwechselt werden. Nur die Reflexion über Sinn und Ziel kann einen Rahmen geben, der verhindert, daß letztlich unbrauchbares entsteht.

Neben den Schwierigkeiten in der Konzeptionsbildung und der damit verbundenen sinnvollen Modularisierung besteht ein zweites Hauptproblem in der unterschiedlichen Entwicklung der einzelnen Teilgebiete. Im Publikationsbereich ist ein großer Rückstand gegenüber den Feldern der Bestandsrepräsentation und der Erschließung im engeren Sinne zu konstatieren. Dies ist wiederum auf das oft fehlende Gesamtkonzept und die Bevorzugung eines schrittweisen Vorgehens mit der Konzentration auf die Probleme der Digitalisierung und der Erschließung im engeren Sinne zurückzuführen.

Dementsprechend ist die erworbene technische und konzeptionelle Kenntnis in diesen beiden Bereichen am höchsten. Die unterschiedlichen Scannertypen, Auflösungsstufen, Farbtiefen oder Dateiformate der Bilddigitalisierung kennen inzwischen die meisten Bibliothekare und Archivare. Auch grundsätzliche Fragen der langfristigen Speicherung sind ihnen wohl vertraut. Für die Metadaten gibt es ebenfalls feste Vorstellungen, wie diese auszusehen haben, daß es gut wäre, gemeinsame Standards zu verwenden und man darauf achten sollte, nur solche Software zu verwenden, die eine Konversion der Daten in standardisierte Austauschformate gewährleisten. In der Digitalisierung sieht man eine neue Herausforderung, die auch angenommen wird: die ganze Bandbreite der technischen Möglichkeiten wird ausgelotet und für alle möglichen Anforderungen, die das komplexe und fragile Material zu erschließender Bestände stellen könnte, wird versucht, eine adäquate Lösung zu finden. In der Erschließung im

engeren Sinne glaubt man sich durch seine scheinbar darauf ausgerichtete Ausbildung und Erfahrung auf der sicheren Seite. Wer sonst, wenn nicht Bibliothekare, Archivare oder andere Fachwissenschaftler wären so vertraut mit den Besonderheiten der Bestände und bewaffnet mit einem theoretischen Rüstzeug, es zu strukturieren, zu beschreiben und für einen effizienten Zugriff aufzubereiten? Daß dies teilweise an den sich verändernden Erwartungen der Benutzer vorbeigehen könnte und nicht unabhängig von den Formen gesehen werden kann, in denen das Datenmaterial elektronisch verwaltet und publiziert werden wird - diese Erkenntnis hat sich zwar noch nicht überall durchgesetzt, ergibt sich aber spätestens bei der praktischen Durchführung. Demgegenüber wird die Frage nach der Art und Weise, in welcher der Benutzer schließlich auf das Material zugreifen und mit ihm arbeiten wird, bislang kaum gestellt. Obwohl sie doch das eigentliche Ziel aller Erschließungsvorhaben sein müßte, ist die Benutzerperspektive meistens nachrangig.

Zugleich steht die Publikation am Schluß des Arbeitsprozesses. So wird es leicht, diesen Aspekt zunächst auszublenden oder sogar ganz vom eigentlichen Projekt abzukoppeln. Diese Wertung mag in ihrer Schärfe zu relativieren sein, da die meisten der untersuchten Projekte noch nicht abgeschlossen sind, bzw. sogar gerade erst angefangen haben. Es bleibt deshalb nicht nur zu hoffen, sondern auch zu erwarten, daß sich in diesem Bereich noch sehr viel tun wird. Für den Augenblick ist aber festzuhalten, daß die einzelnen Teilbereiche digitaler Erschließungsarbeiten sehr unterschiedlich fokussiert werden und mit einem höchst ungleichmäßigen Aufwand an konzeptioneller Arbeit, sowie personellen und finanziellen Ressourcen entwickelt werden. In die Digitalisierung und Beschreibung werden teilweise sehr hohe Summen investiert, während für die Publikation fast nie nennenswerte Mittel bereitgestellt werden. Dies ist auch angesichts sinnvoller Modularisierung nicht in jedem Fall nachzuvollziehen, da sich der Nutzen von Digitalisierung und Erschließung im engeren Sinne an der Frage der Veröffentlichung und an der Akzeptanz der Benutzer entscheidet.

Die traditionell schon vielfältigen Aufgaben von Bibliotheken, Archiven und ähnlichen Einrichtungen erstrecken sich von der Sicherung und Lagerung des Materials, über dessen Erschließung bis hin zur Gewährleistung des Zugriffs durch die Benutzer. Diese drei Bereiche bleiben auch in Zukunft bestehen, weiten sich aber aus und erfahren einige Veränderungen in der Art, wie sie bewältigt werden. Insbesondere der öffentliche Zugriff auf erschlossene Bestände wird sich wandeln. Hier führt die weltweite Vernetzung dazu, daß die Informationsangebote nicht mehr lokal sind, sondern global. Das hat unmittelbar zur Folge, daß auch multiple Bestände (wie z.B. gedruckte Werke) nur einmal erschlossen werden müssen. Unter den Bibliotheken kann dies - positiv gesehen - zu einer stärkeren Verteilung von Aufgaben und zu einer Spezialisierung führen, es bringt aber auch einen höheren Konkurrenzdruck mit sich. Digitale Bibliotheken als Erweiterung und Fortsetzung traditioneller Biblio-

theeken machen nur dann Sinn, wenn sie originäre Angebote vorweisen können, die sie selbst erstellen müssen. Hier ist eine Tendenz absehbar, sich entweder auf bereits vorhandene einzigartige Sammlungen zu konzentrieren, oder in fachlicher Auswahl Schwerpunkte zu setzen und zu versuchen, bestimmte Nischen zu besetzen. Bibliotheken, die weder besondere eigene Sammlungen besitzen, noch spezielle Kompetenz für einzelne Themenbereiche ausbilden, werden keine Zukunft als „digitale Bibliotheken“ haben, sondern bei ihren bisherigen Aufgaben stehen bleiben, deren relative Bedeutung in der Informationslandschaft abnimmt.

Deutlich verändern sich die Formen der Bereitstellung digitaler Materialien. Statt eines schlichten Kataloges oder Findmittels wird man hier in Zukunft mehr anbieten müssen. Die Benutzer werden langfristig komplexe Informationssysteme erwarten, die einen komfortablen Zugriff auf verhältnismäßig tief erschlossene und digital repräsentierte Dokumente ermöglichen. Hierin liegt eine der zentralen Verschiebungen im Aufgabenspektrum beständehaltender Institutionen. Zu fragen ist dabei, wie die umfangreicher werdende Erschließungsarbeit organisiert und verteilt wird und ob sie von den Bibliotheken, Archiven und Forschungseinrichtungen in ihrer ganzen Bandbreite eigenständig durchgeführt werden, oder ob hier in starkem Maße kooperative Vorgehensweisen angewendet werden können. In jedem Fall werden die beständehaltenden Institutionen die zentrale Stelle im Erschließungsprozeß bleiben und ihre Bedeutung sogar erhöhen, weil sie über wichtige Kernkompetenzen der Erschließung im engeren Sinne verfügen, als Schnittstelle zwischen den verschiedenen Spezialaufgaben dienen können und mit ihrer institutionellen und personellen Stabilität als einzige in der Lage sind, eine dauerhafte Benutzbarkeit, Pflege und Fortentwicklung digitaler Bestände zu gewährleisten. Sollten die Verlage und andere Publikationsunternehmen auch weiterhin auf ein stärkeres Engagement in diesem Bereich verzichten, wird den Bibliotheken und Archiven eine publizistische Aufgabe zukommen, die das bisherige Maß weit übersteigt. Wäre es grundsätzlich noch vorstellbar, daß diese Institutionen nur eine „Grundversorgung“ mit Informationen bereitstellen und die Verlage komplexere und stärker ausgearbeitete Publikationsformen übernehmen, so deutet doch derzeit einiges darauf hin, daß diese Aufteilung nicht stattfindet und Bibliotheken und Archive in zunehmenden Maße selbst den gesamten Bereich der Veröffentlichung übernehmen müssen.

Die vorliegende Untersuchung ist ein Versuch gewesen, das inzwischen sehr groß gewordene Feld der digitalen Bestanderschließung zu systematisieren, um einen Überblick über die verschiedenen Teilbereiche und Aufgabenstellungen zu ermöglichen und um die Position einzelner Projekte innerhalb der Gesamtentwicklung besser bestimmen zu können. Nur vor dem Hintergrund von allgemeiner Entwicklung und Zielstellung ist es möglich, ein Erschließungsprojekt sinnvoll in einzelne Arbeitsbereiche zu unterteilen und so anzulegen, daß es benutzbar und zukunftssicher wird. Die Bedingungen für

Benutzbarkeit und Zukunftssicherheit sind es auch, denen jetzt eine stärkere Aufmerksamkeit gewidmet werden sollte. Hier ist vor allem die Entwicklung und Durchsetzung gemeinsamer Standards, die schon bei der Planung zu ermöglichende Offenheit für Erweiterungen (tiefere Erschließung) und Konzeptveränderungen (technische Unabhängigkeit von Beschreibung und Publikation) und die stärkere Orientierung am Benutzer zu nennen.

Die ungleiche Entwicklung der einzelnen Arbeitsbereiche erfordert für die nähere Zukunft eine Schwerpunktsetzung im Bereich des öffentlichen Zugangs zu und der Benutzung von digital erschlossenen Beständen. Die Wünsche, Erwartungen, Anforderungen und (sich verändernden) Arbeitsmethoden der Benutzer müssen untersucht werden und zur Grundlage für digitale Publikationsformen werden. Benutzerschnittstellen und -oberflächen müssen stärker als bisher den Interessenten entgegenkommen und nicht nur die Aufgabe erfüllen, das Datenmaterial in irgend einer Form zugreifbar zu machen. An der Akzeptanz durch den Benutzer entscheidet sich der Sinn und Nutzen digitaler Erschließungskampagnen. Wird dies nicht ernst genommen, werden die mit hohem Aufwand gewonnenen Informationssammlungen zu Datenfriedhöfen verkommen und all jene faszinierenden und für die Wissenschaften potentiell so wertvollen Bestände weiter ihren Dornröschenschlaf halten.